Crystallization and structure elucidation of benzoic acid-specific type III polyketide synthases

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades einer

Doktorin der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Ebtisam Khalil

aus Elmareg / Libyen

1. Referent: Professor Dr. Ludger Beerhues
2. Referent: Professor Dr. Wulf Blankenfeldt
eingereicht am: 24.09.2018
mündliche Prüfung (Disputation) am: 10.12.2018

Druckjahr 2019

vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakulät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Tagungsbeitraege

Khalil, E., Lukat, P., Liu, B., Beerhues, L. Crystallization and structure elucidation of plant type III polyketide synthase enzymes. Poster. DPhG Jahrestagung (2016), München.

Khalil, E., Lukat, P., Liu, B., Blankenfeldt, W., Beerhues, L. Crystallization and structure elucidation of biphenyl synthase and benzophenone synthase. Poster. Biomolecules and Nanostructures 6 (2017), Podlesice, Poland.

Acknowledgments

Without **Allah** I would not be here right now in this moment keep writing my PhD thesis. I thank him foremost for all the countless blessing that he provided me.

My beloved country **Libya** which now is suffering from an incomprehensive war, my thankful does not make any sense in comparison to what Libya gave me in my whole life. I hope it will return to what it used to be.

The most important person to be mentioned is my supervisor **Prof. Dr. Ludger Beerhues**. With his continuous support I was always motivated to give my best in this challenging work. I deeply appreciate him for giving me the opportunity to join his workgroup because he was always open to hear my ideas and opinions. Each time I came to him I knew that he will just relieve my straggling thoughts and will only choose what could be the best for me. I am really so grateful for his trust and his patience. Seriously, at this point I cannot find the right words anymore and would briefly say: Thank you a lot for everything.

Another one was always there for me when I needed help; this one in particular is **Dr. Benye Liu**. I want to thank him for his valuable discussion and answering my silly questions in a way that made me each time not hesitating to knock his door.

I may call him the secret weapon of this work, the crystallographer and my teacher **Dr. Peer Lukat**. I could not have any success without him. If someone says I was lucky in this work, it was in fact the working with him. He made my PhD journey easier when I thought it may be impossible. Not to forget the number of facilities and equipments for protein crystallization, which were provided by the working group of **Prof. Dr. Wulf Blankenfeldt** at the Helmholtz Centre for Infection Research (**HZI**), Braunschweig.

Dr. Leif Barleben was the person with whom I started this work. I am really so grateful to him for providing me with the most basic and essential instructions needed to tackle this work in the correct way.

Like being with my family, this feeling was provided by the presence of **Doris Glindemann** and **Ines Rahaus**. I thank them for their care, friendly conversations and the positive energy that made me forget all day work problems.

I deeply appreciate the help of **Dr. Raeiner Lindigkeit**. He used a part of his valuable time to solve all internet and computer problems that we were complaining about.

I heartily appreciate the support of my close friends **Rabeia Abdallah** and **Nargis Elgahmi** at anytime I was sad or happy. It was exciting and fun to enjoy being in Germany with them. I am also thankful to the triple **Dr. Mohammed Nabil**, **Dr. Islam Elawaad** and **Dr. Mina Elnoshey** for the unforgettable memories that I had with them. I thank them for sharing their knowledge and experiences with me in addition to their permanent support.

My gratitude is extended to **Dr. Eman Adelrahman**, **Dr. Ines Belhadj** and **Dr. Sahar Abdelaziz** for the work help in my first days. I appreciate the help of **Dr. Frauke Gumz** and **Dr. Einar Stauber** for solving any problems I had in the incubators, the centrifuges and the gel filtration chromatography.

I am thankful to **Dr. Mariam Gaid** for her continuous listening and advices especially in my personal life. I would like to convey my thanks to **Mohammad Nagia**, **Sara Nassar** and all other colleagues I have worked with for their assistance, friendship, and nice work atmosphere.

I acknowledge the **Libyan Government** for the financial support and I am so grateful to the **Faculty of Pharmacy of Benghazi University** for giving me the chance to study in Germany. I would like to thank all my Professors and colleagues in Benghazi in addition to all of my teachers and friends in my city Elmarj.

Speechless and not finding the correct words to express my thanks and gratefulness to my dear **Father** and **Mother** for their constant support, prays and love. There is no appropriate reward for their sacrifices but I hope they will be proud of me. Many thanks to my beloved **Brothers** and **Sisters** for their tenderness, kindness and continous care. May God protect and bless them.

Finally, to my sweet family, my soul, my husband **Fakhri** who encouraged me and endured with me all the difficulties that I suffered from in my life. He was my back and secure, without him I could not achieve this work. To my lovely children **Rinad**, **Abdelrahman** and **Rawand**, they brought me the joyfulness and happiness and reminded me each day that life does not deserve to be worried about. May God keep them all healthy and safe.

Ebtisam Khalil
Braunschweig, Germany

Contents

List of abbreviations

Enzymes

AaOKS	*Aloe arborescence* octaketide synthase
AaPCS	*Aloe arborescence* pentaketide chromone synthase
AhSTS	*Arachis hypogea* stilbene synthase
AmQNS	*Aegle marmelos* quinolone synthase
CeBPS	*Centaurium erythraea* benzophenone synthase
ClCURS	*Curcuma longa* curcumin synthase
CmACS	*Citrus microcarpa* acridone synthase
CmQNS	*Citrus microcarpa* quinolone synthase
OsCUS	*Oryza sativa* curcuminoid synthase
FAS	Fatty acid synthase
FhCHS	*Freesia hybrida* chalcone synthase
GhPS	*Gerbera hybrida* 2- pyrone synthase
GmBPS	*Garcinia mangostana* benzophenone synthase
HaBPS	*Hypericum androsemum* benzophenone synthase
HaBPST135L	*Hypericum androsemum* BPST135L
HsBPS	*Hypericum sampsonii* benzophenone synthase
HsBPST135I	*Hypericum sampsonii* BPST135I
HsBPST135K	*Hypericum sampsonii* BPST135K
HuPKS1	*Huperzia serrata* polykeide synthase 1
KAS	Ketoacetyl synthase
MdBIS	*Malus domestica* biphenyl synthase
MsCHS	*Medicago sativa* chalcone synthase
PKS	Polyketide synthase
PsSTS	*Pinus sylvestris* stilbene synthase
RpALS	*Rheum palmatum* aleosone synthase
RpBAS	*Rheum palmatum* benzalacetone synthase
SaBIS	*Sorpus aucuperia* biphenyl synthase
THNS	1, 3, 6, 8-tetrahydroxynaphthalene synthase

Nucleotides

A	Adenine
C	Cytosine
G	Guanine
T	Thymine

Amino acids

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

Other abbreviations

°C	Degree Celsius
µl	Microliter
µM	Micromolar
APS	Ammonium peroxydisulfate
bp	Base pair
BNY	Bis-noryangonin
BSA	Bovine serum albumin
cDNA	Complementary deoxyribonucleic acid
CoA	Coenzyme A
CTAL	*P*-Coumaroyltriacetic acid lactone
d	Days
DAD	Diode array detector
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DTT	1, 4-Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
Fig	Figure
FPLC	Fast protein liquid chromatography
g	Gram
h	Hour

HPLC	High performance liquid chromatography
IEC	Ion exchange chromatography
IMAC	Immobilized metal ion chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
ITC	Isothermal calorimetry
kDa	Kilodalton
LB	Luria broth
M	Molar
min	Minute
ml	Milliliter
mM	Millimolar
MR	Molecular replacement
MST	Microscale thermophoresis
ng	Nanogram
Ni-NTA	Nickel-nitrilotriacetic acid
nm	Nanometer
No	Number
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
rpm	Revolution per minute
R.T	Room temperature
s	Second
SDM	Site-directed mutagenesis
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SOC	Super optimal broth with catabolite repression
TAE	Tris-acetate-EDTA
TAL	Triacetic acid lactone
TCEP	Tris (2-carboxyethyl) phosphine
TEMED	N, N, N, N`-tetramethylethylenediamine
Tm	Melting temperature (primer)
Tris	Tris (hydroxymethyl) aminomethane
UV	Ultraviolet
w	Week
WT	Wild type

List of Figures

List of Tabels

I. Introduction

Among the huge number of natural products distributed all over the world, polyketides constitute one of the most important families because of their broad biological activities, such as antiviral (Takasaki et al., 1990), antibacterial (Yamakoshi et al., 1992) and anticancer (Yin et al., 2007) properties, which make them good candidates for new drug discovery. They already enfold serious classes of medicaments, such as picromycine, the first isolated antibiotic from the bacterium *Streptomyces venezuelae* (Suhren, 1951), and aspergiolides A, B from the fungus *Aspergillus glaucus*, which have cytotoxic effects (Du et al., 2008). In plants, curcuminoids (curcumin) and flavonoids have anti-inflammatory (Bisht et al., 2010) and anti-cancer effects (Khan et al., 2008). Another example is the eucalypt polyketides, which have antiviral, antibacterial and anticancer effects. The plant is used by the Australians in the treatment of influenza, fever, toothache, colds and snakebites (Ghisalberti, 1996).

Polyketides are widely distributed among prokaryotic and eukaryotic organisms. They are named polyketides because they all arise from polyketo chains. They are biosynthesized through a serious of decarboxylative condensation reactions catalyzed by a group of enzymes called polyketide synthases (PKSs), which divide further into three main groups: Type I, Type II and Type III PKSs (Hopwood, 1997).

Type III PKSs have, compared to the other types, drawn much attention, as they are smaller and simpler but still maintain the catalytic function of polyketide chain elongation and cyclization. In 1999, chalcone synthase (CHS) from *Medicago sativa* (alfalfa) was the first enzyme of this class that has been crystallized and still is the best studied until now (Ferrer et al., 1999). Recently, many new CHS-like enzymes have been crystallized, which showed a structural basis for the functional diversity of these enzymes. In this study, crystallization of two enzymes of the same family, benzophenone synthase (BPS) and biphenyl synthase (BIS), was carried out.

BPS was cloned from cell cultures of different *Hypericum* species like *H. androsaemum* and *H. sampsonii* (Liu et al., 2003). BIS was cloned from cell cultures of *Sorpus aucuparia* and *Malus domestica* (Liu et al., 2007). Both enzymes were expressed successfully in our laboratory and subjected to extensive characterization. They utilize the same starter substrate, benzoyl-CoA, which is a rare starter substrate for type III PKSs. BPS forms 2,4,6-trihydroxybenzophenone via the Claisen condensation mechanism, but BIS forms 3,5-dihydroxybiphenyl using the aldol condensation mechanism (Liu et al., 2003, Liu et al., 2007). These variations in the cyclization mechanism leading to product variability despite the same starter substrate made both enzymes good candidates for crystallization and structure elucidation.

1. Polyketide synthases (PKSs)

They catalyze the biosynthesis of polyketides using an acyl-CoA molecule followed by subsequent elongations to produce the final chain in a similar pattern to fatty acid synthases (FASs).

1.1. Major differences between PKSs and FASs:

***Substrate variation*:** PKSs use different starter substrates and thus exhibit a wide range of substrate tolerance, whereas FASs use only acetyl-CoA as starter substrate for the synthesis of saturated and non-saturated fatty acids. Acetyl-CoA is carboxylated to yield malonyl-CoA, which is used by both enzymes as chain extender.

***Post-PKS modification*:** In contrast to fatty acids, further changes can occur in the structure of the released polyketide chains, for example oxidation to introduce hydroxyl or carbonyl groups and methylation at oxygen, nitrogen or carbon centers.

These differences make the polyketides a promising source of natural products biosynthesis (Revill et al., 1996) (Staunton & Weissman, 2001).

2. Types of PKSs

2.1. Type I PKSs

They are long polypeptides consisting of many catalytic components that include three essential domains: β-ketosynthase (KS), acyl transferase (AT), and acyl carrier protein (ACP). Other domains may be added, such as dehydratase (DH), enoylreductase (ER), β-ketoreductase (KR), and thioesterase (TE). These components are linked to each other by covalent bonds in a similar pattern of type I fatty acid synthases (Figure I.1). An example of this group is 6-deoxyerythromycine B synthase (DEBS), which synthesizes the macrolide antibiotics such as erythromycin A (Smith & Tsai, 2007). They are either modular (each active site is used only once) or iterative (the active site is re-used repeatedly).

2.2. Type II PKSs

They are multienzyme complexes, each with a single active site which can be used only once or iteratively (Figure I.1). They resemble ß-ketoacyl synthase II of type II FAS and biosynthesize most of the aromatic polyketides (Watanabe & Ebizuka, 2004); an example is tetracenomycine PKS that synthesizes the tetracenomycine c antibiotic (Shen, 2003).

2.3. Type III PKSs

In comparison to the other types of PKSs, this class in particular has more intrinsic characters and biosynthesis steps, which are the starter substrate preference for reaction initiation, the number of acetyl molecules added for chain elongation, and the cyclization pattern in addition to reaction termination (Austin & Noel, 2003).

Mostly, they are called plant PKSs because of their abundance in plants. They are homodimeric proteins, in which the active site of each monomer catalyzes iterative reactions to produce the polykeide chain. They are simpler in structure in comparison to the other types (Figure I.1) (Yu et al., 2012).

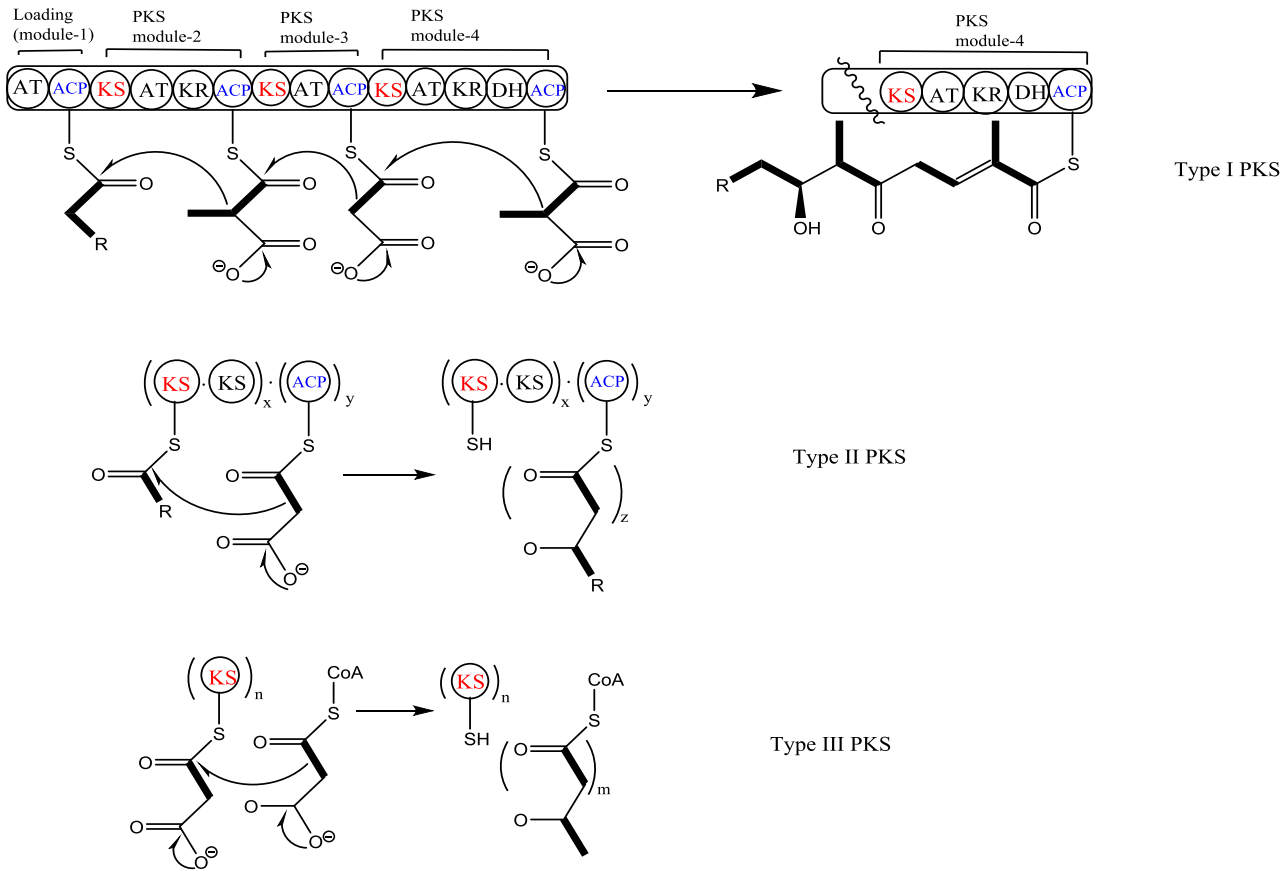


Fig I.1 The different classes of polyketide synthase enzymes.

2.3.1. Type III PKSs division

Only a small modification of the active-site architecture leads to a significant functional diversity of type III PKS enzymes. In general, it is the size and the shape of the active-site cavity that determine the starter molecule selectivity, polyketide chain length and cyclization pattern of the final product (Li et al., 2011).

According to the reaction and cyclization mechanisms, type III PKSs are grouped as follows.

Chalcone synthase-type (CHS), which utilizes the Claisen condensation mechanism, in which the intramolecular cyclization is from C6 to C1 (Figure I.2). This mechanism is also used for fatty acids biosynthesis.

Stilbene synthase-type (STS), which utilizes the aldol condensation mechanism, in which the intramolecular cyclization is from C2 to C7 with loss of C1 as CO_2 (Figure I.2).

Coumaroyl triacetic acid synthase-type (CTAS), which utilizes lactonization, in which a lactone forms between oxygen of C5 to C1 (Figure I.2).

Stilbene carboxylate synthase-type (STCS), whih utilizes a mechanism similar to aldol condensation but without decarboxylation of C1 (Figure I.2).

Other types can undergo condensation but without cyclization, such as benzalacetone synthase **(BAS)** from *Rheum palmatum* and curcuminoid synthase **(CUS)** from *Oryza sativa* (Flores-Sanchez & Verpoorte, 2009). Below, various type III PKSs are presented in more detail, with a focus on their crystal structures.

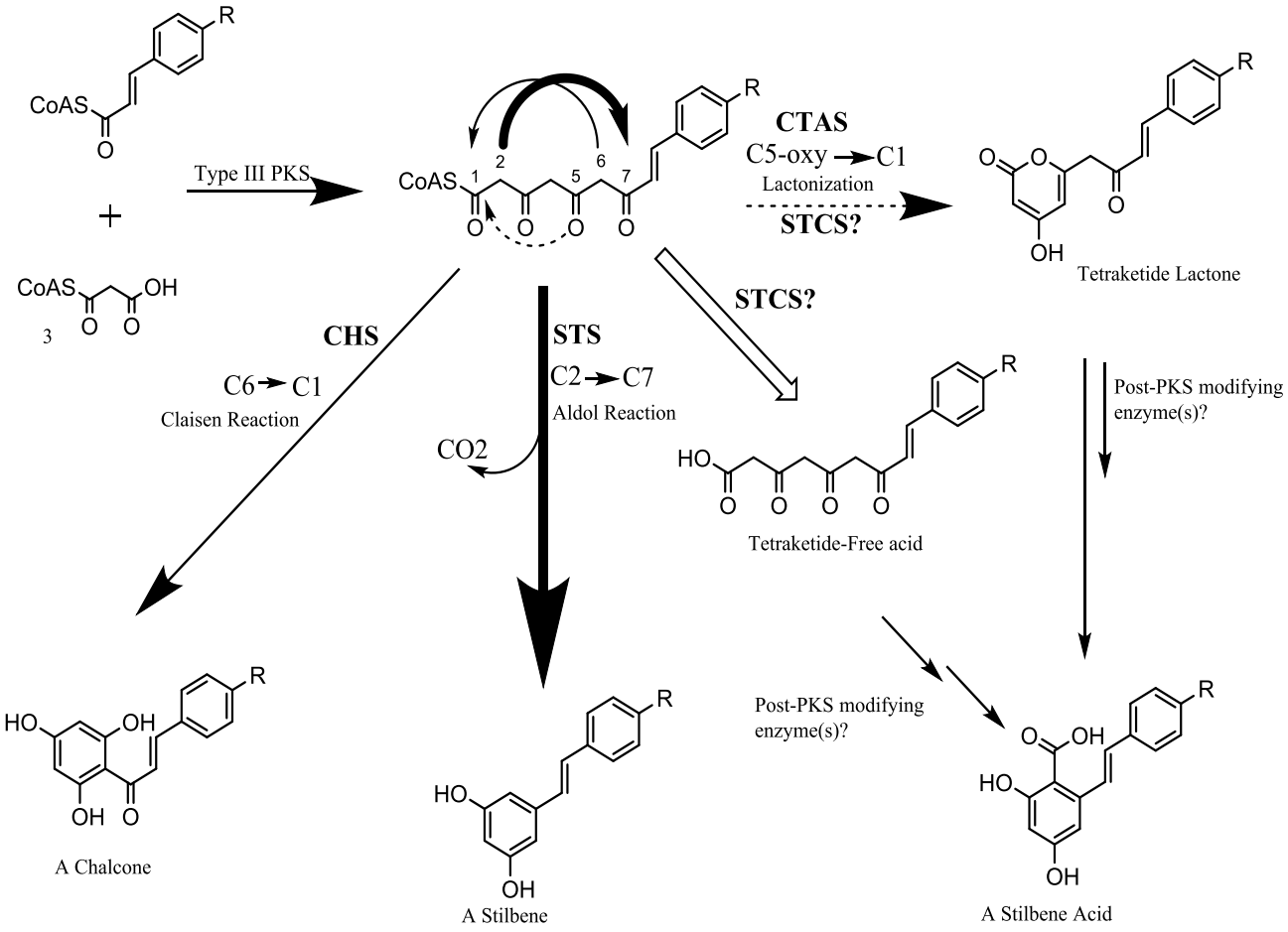


Fig I.2 The different reaction patterns of type III polyketide synthases.

2.3.2. Plant type III PKSs

2.3.2.1. Chalcon synthase (CHS)

CHS is the best studied enzyme of type III PKSs. The first isolation of a *CHS* gene from parsley (*Petroselinum hortense*) was in 1983 (Reimold et al. 1983). This enzyme catalyzes the decarboxylative condensation of p-coumaroyl-CoA with three molecules of malonyl-CoA to produce naringenin chalcone, which is the precursor for flavonoid biosynthesis (Fig I.3). Flavonoids are an important class of natural products, which have many medicinal effects such as antioxidant, anticancer, antiasthmatic and antimitotic properties (Austin & Noel, 2003). The importance of the CHS enzyme increased even more in 1999 as the first crystal structure of CHS2 from the legume *Medicago sativa* (alfalfa) (MsCHS2) was published, which is the first crystallized enzyme of type III PKSs (Ferrer et al., 1999).

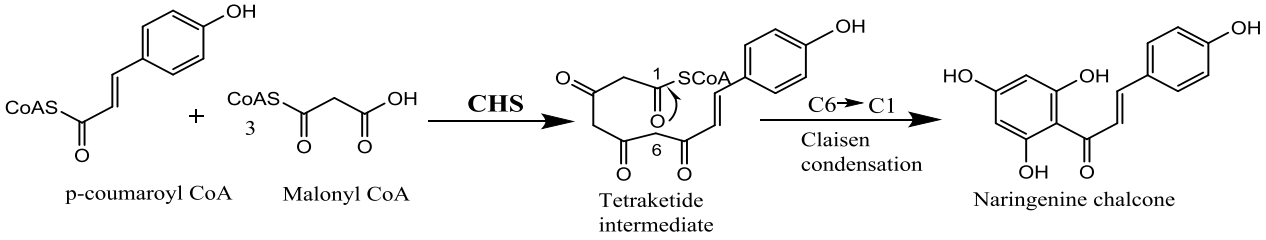


Fig I.3 The reaction catalyzed by chalcone synthase.

According to its structural architecture, MsCHS2 was grouped together with ketoacetyl synthases (KAS III) to be enfolded in thiolase-fold enzymes. They share common features starting from αβαβα core, dimerization, active site location and the catalytic cysteine residue for covalent binding of substrate and intermediates. The MsCHS2 homodimer contains two active site cavities located at the bottom of each monomer. At the dimer interface a loop of six residues protrudes from each monomer starting from Thr132 and ending with Pro138. Only Met137 of this loop contributes to the opposing monomer's active site cavity (Figure I.4) (Austin & Noel, 2003).

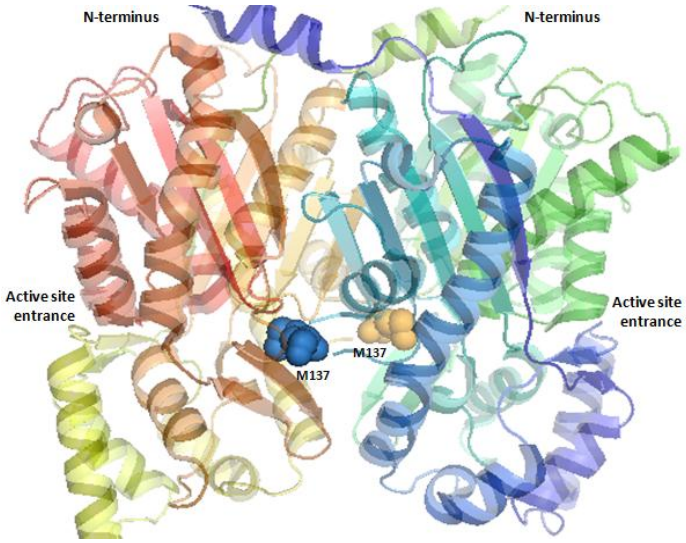


Fig I.4 The crystal structure of *Medicago sativa* chalcone synthase 2.

The crystal structure of MsCHS2 revealed essential residues for an active enzyme, which are Cys164, His303 and Asn336 (**catalytic triad**) (Figure I.5). Among them Cys164 is in the top, its absence will abolish the enzyme activity (Lanz et al., 1991). The importance of Cys164 appears from its nucleophilic character that provides an efficient binding site for polyketide intermediates. Mutation of Cys164 to Ser or Ala reduces but maintains the activity of decarboxylation of malonyl-CoA (Jez et al., 2000b).

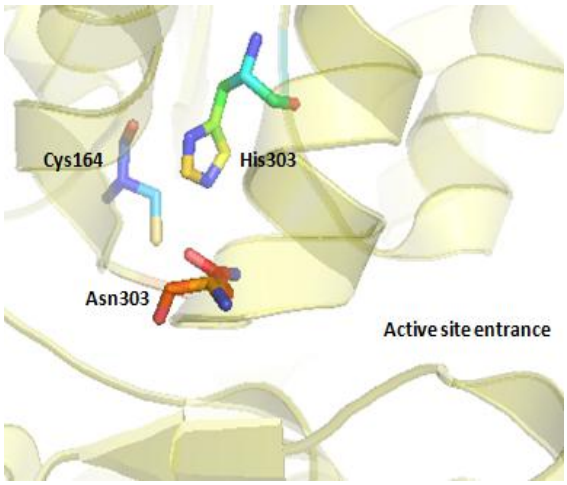


Fig I.5 The catalytic triad of *Medicago sativa* chalcone synthase 2.

The structure also illustrated the **CoA tunnel** that connects the embedded active site cavity with the surrounding solvent. The binding of CoA within this area allows the positioning of the substrate toward the active site. The residues lining this tunnel are Lys 55, Arg 58, Lys 62 and Ala 308, each residue binding with CoA of the corresponding substrates. The two lysines and the argenine form hydrogen bonds with two phosphates of the CoA and the amide nitrogen of the alanine forms a hydrogen bond with the first carbonyl of the pantetheine moiety. The pantetheine arm is a flexible important part of the CoA as it expands through the tunnel and helps in the positioning of the terminal substrate bound to the thioester linkage near the Cys164 (Figure I.6) (Ferrer et al., 1999)

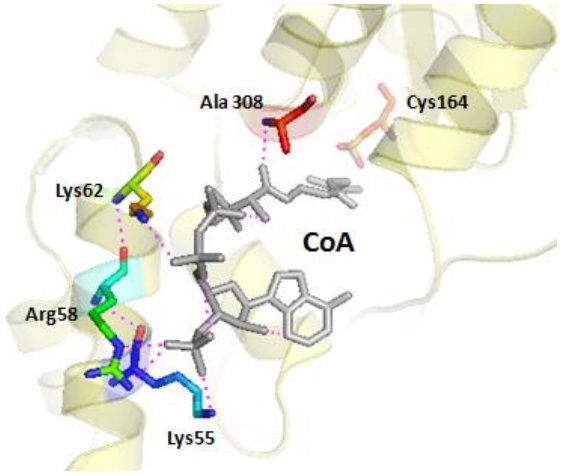


Fig I.6 The CoA binding tunnel of *Medicago sativa* chalcone synthase 2.

The **coumaroyl-binding pocket** is located at the lower left side of the CoA-binding tunnel containing Ser 133, Glu 192, Thr 194, Thr 197 and Ser 338). These residues surround the substrate and make the binding mostly through van der Waals forces (Figure I.7) (Ferrer et al., 1999). The steric effect of these residues participates in the choice of the different substrates that will be accepted by type III PKSs. Only Glu192 shows a high conservation in comparison to the other residues. In 2004, a new substrate binding tunnel was discovered in a type III PKS from *Mycobacterium tuberculosis*, namely PKS18. This enzyme showed a unique broad specificity for aliphatic long chain CoAs (C_2-C_{20}) to produce tri-and tetraketide pyrones. In comparison to the MsCHS2 tunnel (16 Å), this tunnel is bigger in size covering about 20 Å. This size difference influencing the substrates specificity emphasizes the importance of this tunnel in formation of natural product diversity (Sankaranarayanan et al., 2004).

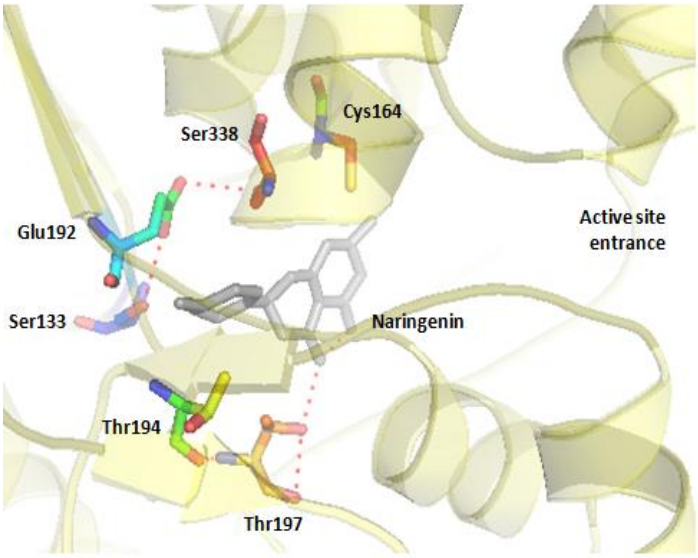


Fig I.7 The substrate binding pocket of *Medicago sativa* chalcone synthase 2.

The **malonyl-binding pocket** covers a large cavity adjacent to Cys 164, including Thr 132, Met 137, Phe 215, Ile 254, Gly 256, Phe 265 and Pro 375. Van der Waals force also dominates the interaction in this area as these amino acids provide few hydrogen bond donors. The size of this pocket controls the number of the acetyl molecules added to the starter substrate and the length of the growing polyketide intermediate. The shape and the surface geometry also control the cyclization and the aromatization into a new ring system (Figure I.7) (Ferrer et al., 1999). Further studies proved the participation of the G372FGPG loop in the end result of the cyclization reactions. Pro-375 in this loop is strictly conserved in all other members of the CHS superfamily cloned so far. S397FGFG is another similar loop that was found in KAS II of *E. coli* and it acts as the flexible gate to the active site. Such similarity and conservation of important residues between CHS and KAS in polyketide and fatty acid synthases, respectively, strike the light on the evolution of seconderay metabolism enzymes, as CHS, from enzymes that are involved in the primary metabolism to meet the need for production of various secondary metabolites, such as flavonoids (Figure I.8) (Suh et al., 2000).

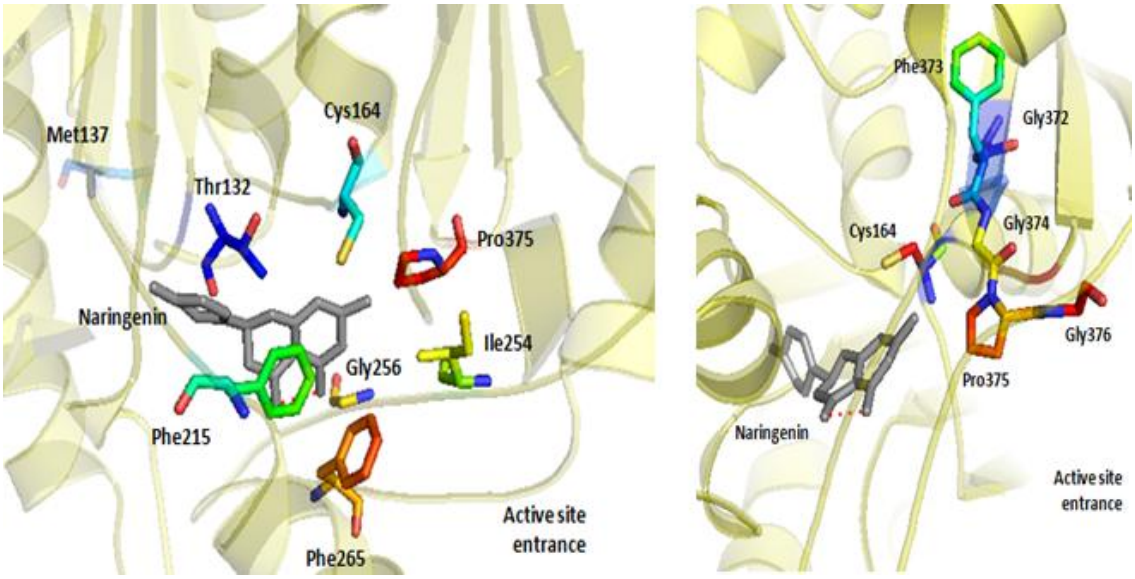


Fig I.8 The cyclization and elongation pockets of *Medicago sativa* chalcone synthase 2. The G372FGPG loop is shown on the right.

2.3.2.2. Stilbene synthase (STS)

As STS is less abundant but not less important than CHS, this enzyme produces resveratrol (3,5,4'-trihydroxy-*trans*-stilbene), which is considered as one of the compounds that are involved in the health benefits associated with red wine and grape juice consumption. These benefits are linked to the activities that resveratrol has, such as antitumor, anticholesterol, anti-inflammatory and cell death reduction properties (Romero-Perez et al., 1999). Over their approx. 400 residues, CHS and STS share 75%–90% amino acid identity. Both enzymes catalyze iterative condensation of three molecules of malonyl-CoA and one molecule of p-coumaroyl-CoA, resulting in a tetraketide intermediate. The cyclization pattern of this intermediate differs between the enzymes. In case of CHS, C6→C1intramolecular cyclization and cleavage of the C1 thioester linkage will occur to produce naringenin chalcone. However, the intermediate undergoes C2→C7 cyclization by removing of C1 in STS. This step requires thioesteras-like hydrolysis of the thioester linkage to aid in the decarboxylative elimination of C1 carboxylate (Figure I.9).

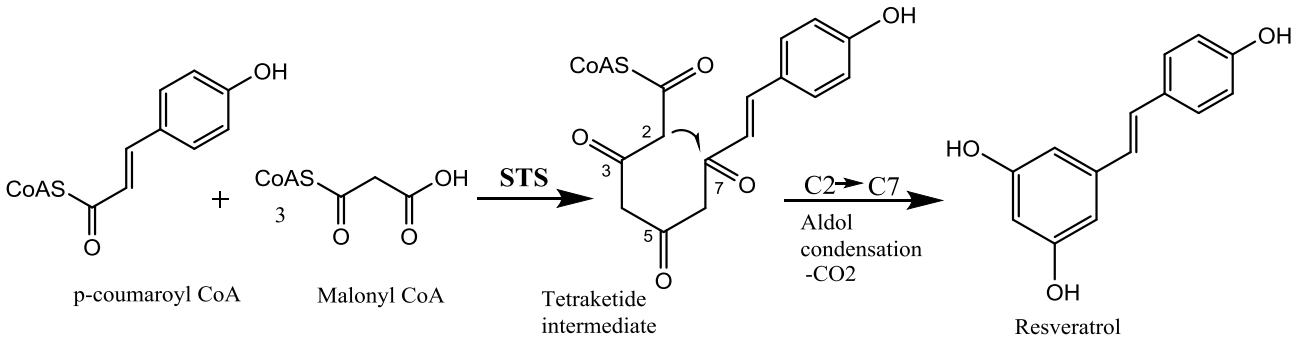


Fig I.9 The reaction catalyzed by stilbene synthase.

It was possible to crystallize the first STS in 2004. It was also generated by mutation from alfalfa CHS2 and it showed important residues responsible for the aldol cyclization specificity. Such results were supported when the first crystal structure of STS from *Pinus sylvestris* (PsSTS) was solved (Austin et al., 2004a). The active site cavities were similar in CHS and STS but PsSTS contained two major different regions, namely regions A and B. Region A is located at the dimer interface and divided into three areas. Area 2 presents a six residue loop starting from Thr135 until Leu140. Region B presents area 4, which consists of a three residue loop located on the outer surface of the CoA binding tunnel. To understand the different reaction mechanisms, those two different regions were subjected to mutation in alfalfa CHS2. By substitution of residues in these regions to the corresponding ones of PsSTS, a creation of a new enzyme (18xCHS) was achieved, which produced resveratrol rather than chalcone as the major product. Region A was more important than region B for driving the aldol cyclization mechanism. Deeper studying of this region proved that the area 2 loop is the critical location for the cyclization. In MsCHS2 three residues in this area were in contact with the active site cavity. Thr132 and Ser133 cover their own monomer active site cavity and Met137 is at the active site cavity of the other monomer (Figure I.10).

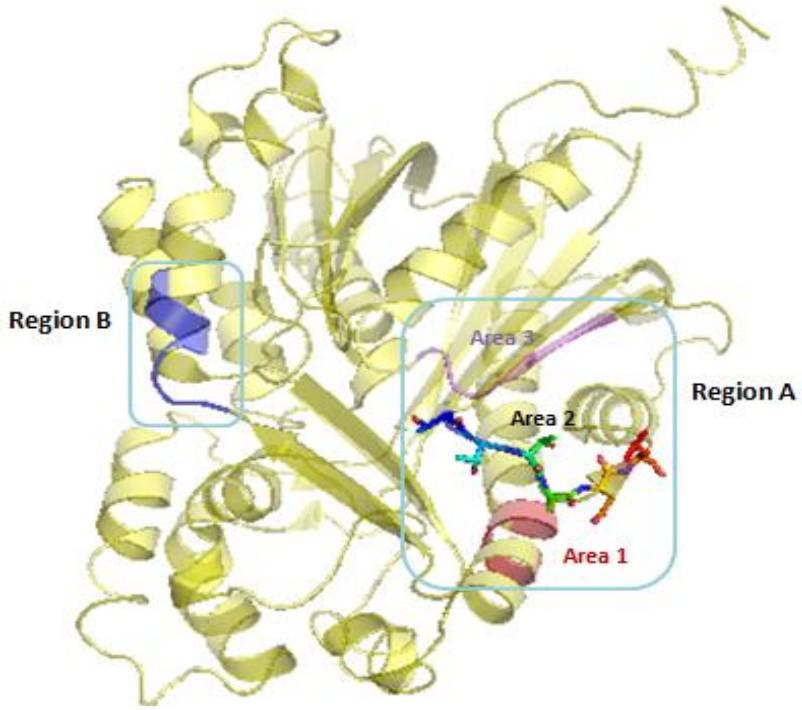


Fig I.10 The important regions in the crystal structure of *Pinus sylvestris* STS. Region A comprises area 1 (red), area 2 (colored), and area 3 (purple). Region B is blue.

Residue 137 was replaced before with the corresponding one of PsSTS (Met in MsCHS2, Leu in PsSTS) but there was no effect on the activity. Substitution of residue 133 (Ser in MsCHS2 and Thr in PsSTS) had a drastic effect on the conformation of the area 2 loop, but it is far from the active site and it is unreasonable to be involved in the aldol cyclization mechanism. Only when Thr132 was substituted with Ala in the 18xCHS mutant, the production of chalcone incrased over stilbene (Austin et al., 2004a).

2.3.2.3. Benzoic acid specific type III PKSs

Two important enzymes were cloned in our group, which follow the same reaction mechanisms of CHS and STS, as follows.

2.3.2.3.1. Benzophenone synthase

This enzyme catalyzes the formation of the C_{13} skeleton of important secondary metabolites, for example the polyprenylated benzophenones such as guttiferone F with anti-HIV and antimicrobial effects. Benzophenones are also the precursor for 1,3,5 and 1,3,7–trihydroxyxanthones, which posses antitumor and antibacterial activities, respectively. BPS condenses one molecule of benzoyl CoA and three molecules of malonyl-CoA to produce the linear tetraketide intermediate, which undergoes the cyclization mechanism of CHS using Claisen condensation to end up with 2,4,6-trihydroxybenzophenone (Figure I.11) (Liu et al., 2003).

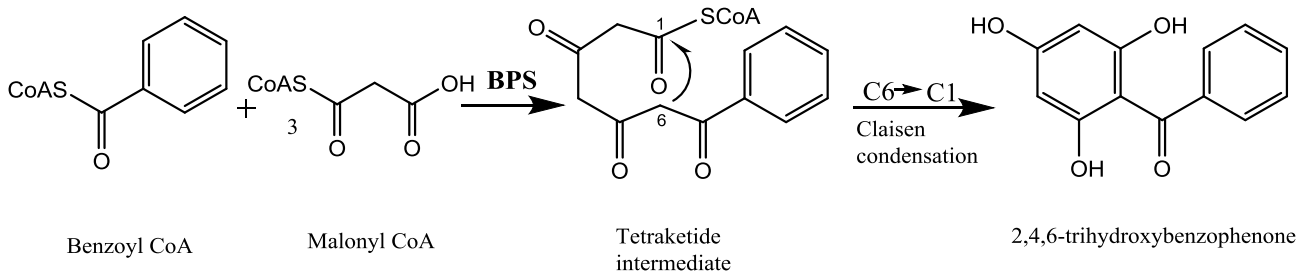


Fig I.11 The reaction catalyzed by benzophenone synthase.

Several BPSs were cloned in our work group and had different characteristics, as follows.

A. Benzophenone synthases from *Hypericum* species

Hypericaceae known previously as Guttiferae or Clusiaceae (Stevens, 2007) is a large family in the plant taxonomy consisting of 500 species distributed all over the world (Sanmartín, 2012). They are specialized by the presence of different important classes of secondary metabolites consisting mainly of naphthadianthrones (hypericin, pseudohypericin), phloroglucinols, flavonoids, phenylpropanoids and xanthones (Bruni et al., 2005). Leave extract of *Hypericum perforatum* has been used for many years in Europe to treat mild to moderate depression (Crockett & Robson, 2011). Other important activities such as antiviral and anti-inflammatory effects (Birt et al., 2009) made the genus *Hypericum* a good candidate for research studies due to the important economical value of this plant in the herbal remedy industry.

***Hypericum androsaemum* BPS (HaBPS)** is the first BPS that was cloned from the cell culture of this species, which is distributed mainly in Western and Southern Europe (Rosaria Perrone, 2013). It has about 59.1% amino acids sequence identity to MsCHS2 consisting of 395 amino acids and has a molecular mass of 42.8 kDa. It shows a maximum activity with benzoyl-CoA in comparison to 3-hydroxybenzoyl-CoA, which is the second preferred substrate for this enzyme. Many mutational works were done on this enzyme in particular because of its low Km values that reach to 5.7µM for benzoyl-CoA and 23.1 µM for malonyl-CoA (Liu et al., 2003). Depending on the structure of MsCHS2 there were 7 amino acids in the initiation/elongation cavity different from HaBPS, but mutation of these residues to the corresponding ones of CHS failed to convert BPS into functional CHS. On the other hand, a triple mutation of HaCHS could produce a functional enzyme that accepted benzoyl-CoA over p-coumaroyl-CoA, but the efficiency of this mutant was 10-fold less than that of the wild type BPS. The most successful mutation was a single amino acid substitution of threonine at position 135 to leucine, which converted HaBPS into a functional 2-pyrone synthase (HaBPST135L).

The **HaBPST135L** mutant accepts also benzoyl-CoA as a substrate but it utilizes only two molecules of malonyl-CoA to produce 6-phenyl-4-hydroxy-2-pyrone as a major product and 2, 4, 6-trihydroxybenzophenone as a minor product, which is the reverse situation for the wild type HaBPS (Fig I.12). This succefull mutation came up because of the importance of threonine at position 135 in the reaction mechanism in which its hydroxyl group forms hydrogen bonds with other residues in the active site and allows the entrance of the substrates to the elongation pocket. In case of leucin another pocket opens for substrate entrance which, however, can accommodates only the triketide intermediate after condensation of only two molecules of malonyl-CoA. This mutant also reacts with 3-hydroxybenzoyl-CoA and in this case the hydroxyl group of the substrate blocks the entrance even to the new pocket, which can be explained by the inactivity of HaBPST135L with 3-hydroxybenzoyl-CoA (Klundt et al., 2009).

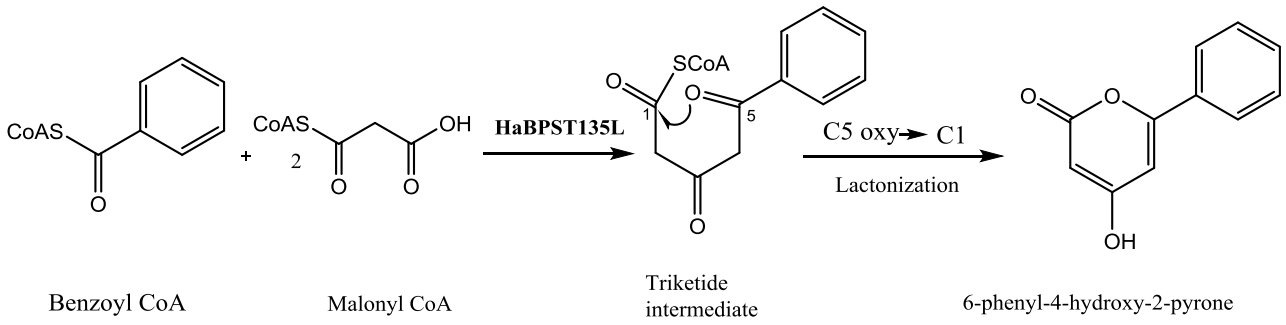


Fig I.12 The reaction catalyzed by *Hypericum androsemum* BPST135L.

Beside HaBPS, two other BPSs were cloned from *Hypericum* species; one from *Hypericum perforatum* (HpBPS) and the other from *Hypericum calycimum* (HcBPS). They are quite similar to HaBPS but mutational experiments on those enzymes were less characteristic in comparison to HsBPS, which was cloned from *Hypericum sampsonii* (Abdelaziz, 2014).

***Hypericum sampsonii* BPS (HsBPS)** is a newly cloned enzyme from the herbal medicinal plant *Hypericum sampsonii*, which is used in China for the treatment of burns, snakebites, and diarrhea and also in Taiwan for treating blood stasis and hepatitis. These effects are associated with the presence of a lot of bioactive polyprenylated benzoylphloroglucinol derivatives, such as sampsoniones (Xiao et al., 2007, Chen et al., 2014). HsBPS prefers benzoyl-CoA as HaBPS to produce 2,4,6-trihydroxybenzophenone but in contrast it accepts also salicoyl-CoA (Huang et al., 2012). At the amino acids level, HsBPS is quite similar to HaBPS characterized by only six different amino acids that lie on the surface and far away from the active site cavity. They are Asp54, Ala146, Gly230, Ala235, Leu359 and Ser360. When any of them were changed from HaBPS to the corresponding one of HsBPS, the enzymatic activity was still the same but the catalytic constant changed. Similar mutation experiments at site 135 were done for HsBPS, yielding two new functional enzymes named HsBPST135I and HsBPST135L, which have the activity of 2-pyrone synthase, and HsBPST135K, which has the activity of hydroxycoumarin synthase (Abdelaziz, 2014).

HsBPST135I is an interesting mutant like HaBPST135L. Although leucine and isoleucine are not much different from each other, the isoleucine mutant converted only HsBPS, but not HaBPS, into functional 2-pyrone synthase. The mutant condenses only two molecules of malonyl-CoA with benzoyl-CoA producing a lactone, which is similar to HaBPST135L and HsBPST135L. When the same mutation experiment was constructed on HaBPS, the produced mutant HaBPST135I did not make any change in the wild type but it gave inactive protein (Abdelaziz, 2014).

HsBPST135K is another interesting single mutant that could convert HsBPS into functional hydroxycoumarin synthase (HCS), which prefers 2-hydroxybenzoyl-CoA (salicoyl-CoA) as a substrate and produces 4-hydroxycoumarin as the main product after the condensation of only one molecule of malonyl-CoA (Fig I.13) (Abdelaziz, 2014). Coumarins rise in plants upon cell damage in which *trans-ortho*-coumaric acid is converted by fungi to 4-hydroxycoumarin that combine with formaldehyde to give dicoumarol. This class of compounds is important for the pharmacology because it belongs to the medicinal anticoagulant drugs (Liu et al., 2010). The mutation of threonine at position 135 into lysine was also succeful in case of HaBPS but the activity of the produced enzyme was much lower in comparison to HsBPST135K (Abdelaziz, 2014). In general, the variation in substrate acceptance leading to product variation is a common character of type III PKSs, as exemplified by 2-hydroxybenzoyl-CoA which led to the formation of coumarins when elicitor-treated cell cultures of *Sorbus aucuparia* were fed with salicoyl-N-acetylcysteamine. The production of coumarins in this case was due to the action of another type III PKS, named biphenyl synthase (Liu et al., 2010).

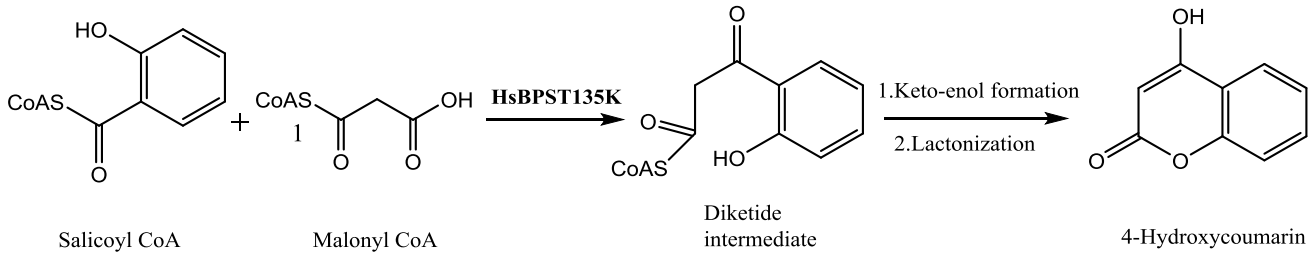


Fig I.13 The reaction catalyzed by *Hypericum sampsonii* BPST135K.

B. *Centaurium erythraea* BPS (CeBPS)

This is another interesting enzyme among the group of BPSs because it prefers 3-hydroxybenzoyl-CoA over benzoyl-CoA and form 2,3',4,6-tetrahydroxybenzophenone, which is converted further to 1,3,5-trihydroxyxanthone by xanthone synthase. *C. etythraea* belongs to the family Gentianaceae and it is known to have a high content of xanthones that play an important role in health benefits (Figure I.14) (Beerhues, 1996).

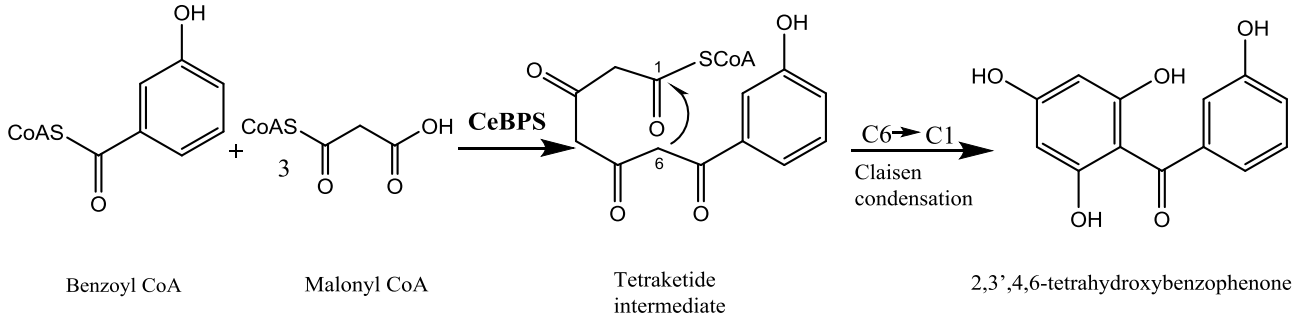


Fig I.14 The reaction catalyzed by *Centaurium erythraea* benzophenone synthase.

C. *Garcinia mangostana* BPS (GmBPS)

This plant belongs to the family Hypericaceae (Guttiferae) and was used as a traditional medicine for topical infections and wound healing. GmBPS prefers the starter substrate benzoyl-CoA as other BPSs to yield the major product 2, 4, 6-trihydroxybenzophenone and it showed 77-78% amino acid sequence identity with other BPSs. At the active site, it has the same amino acids of HaBPS, which differ from the residues in MsCHS2 (V196, L214, G216, G256, L263, F265 and S338). The enzyme showed wide substrate acceptance. From, salicoyl-CoA it produced 4-hydroxy coumarin similar to biphenyl synthase (BIS) from *Sorbus aucuparia*. In contrast to HaBPS, GmBPS accepted the larger aromatic substrates like cinnamoyl-CoA and 4-coumaroyl-CoA but allowed the condensation of two molecules of malonyl-CoA to produce the corresponding triketide lactones (Figure I.15) (Nualkaew et al., 2012).

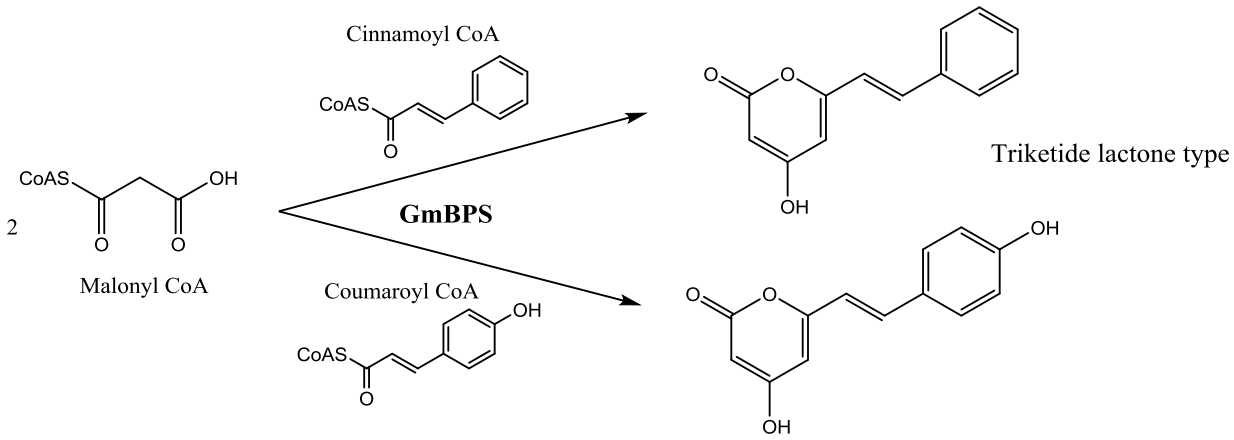


Fig I.15 The reactions of *Garcinia mangostana* benzophenone synthase with larger aromatic substrates.

2.3.2.3.2. Biphenyl synthase

It is the key enzyme responsible for the synthesis of the C_{12} skeleton of biphenyls. Biphenyls are phytoalexins found together with dibenzofurans in large amounts in Maloideae, a subfamily of the Rosaceae which includes economically important fruit trees, such as apple (*Malus*) and pear (*Pyrus*). The accumulation of these phytoalexin increases mostly after fungal infection and for some species after the treatment of heavy metal ions. Such defense mechanisms can help the plant to survive and resist the diseases caused by some pathogens. Aucuparin is the most well known biphenyl, which was first isolated from *Sorbus aucuparia* in 1963.

A. *Sorbus aucuparia* BIS1 (SaBIS1)

It was cloned for the first time in our research group after the observation of high amount accumulation of aucuparin in the cell culture upon treatment with yeast extract. This enzyme also prefers benzoyl-CoA as a substrate and produces 3,5 dihydroxybiphenyl after condensation of three molecules of malonyl-CoA using the aldol condensation mechanism for the cyclization of the linear tetraketide intermediate in a similar manner of STS. It shares about 60% amino acids sequence identity with the other enzymes of the type III PKSs superfamily and has a molecular mass of 43 kDa covered by 390 amino acids. It was grouped closely to BPS in a phylogenetic tree (Figure I.16) (Liu et al., 2004, Liu et al., 2007).

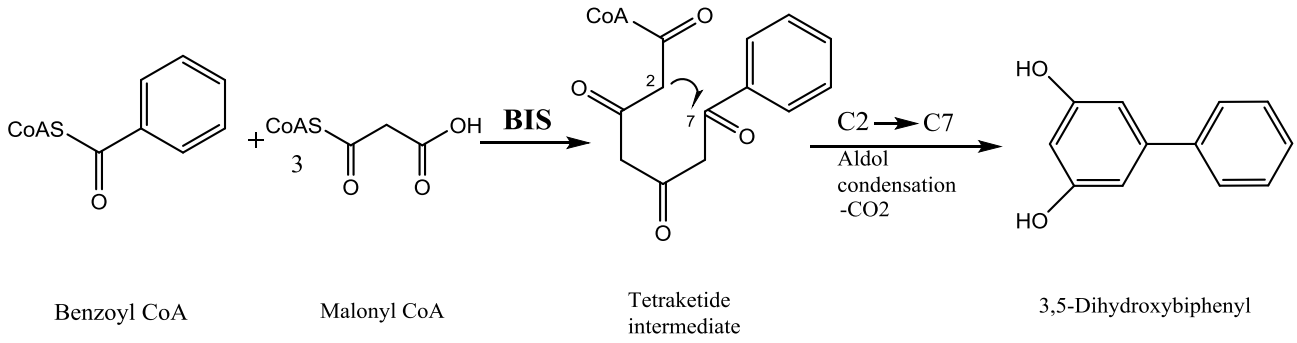


Fig I.16 The reaction catalyzed by biphenyl synthase.

Two other isoenzymes of biphenyl synthase were cloned from the same species namely SaBIS2 and SaBIS3. They share more than 90% amino acid sequence identity with BIS1 but they differ in their preference for ortho-hydroxybenzoyl (salicoyl)-CoA over benzoyl-CoA to yield 4-hydroxycoumarin after a single decarboxylative condensation with malonyl-CoA (Figure I.17) (Liu et al., 2010).

B. *Malus domestica* BIS3 (MdBIS3)

It is quite similar to the previously mentioned biphenyl synthases from *Sorbus aucuparia* and it was cloned together with other isoenzymes (BIS1, BIS2 and BIS4) from fire-blight-infected shoots of apple. Similar to SaBISs, these isoenzymes convert benzoyl-CoA, 2-hydroxybenzoyl-CoA (salicoyl-CoA), and 3-hydroxybenzoyl-CoA to 3,5-dihydroxybiphenyl, 4-hydroxycoumarin, and 6-(3'-hydroxyphenyl)-4-hydroxy-2-pyrone, respectively. Only the *MdBIS3* gene showed a higher expression level in the stems of apple after inoculation with *Erwinia amylovora*, the causative pathogen of the fire-blight disease. Immunofluorescence localization results showed that MdBIS3 localized to the transition zone of the infected stem, which lies between the healthy and the necrotic tissue (Chizzali et al., 2012).

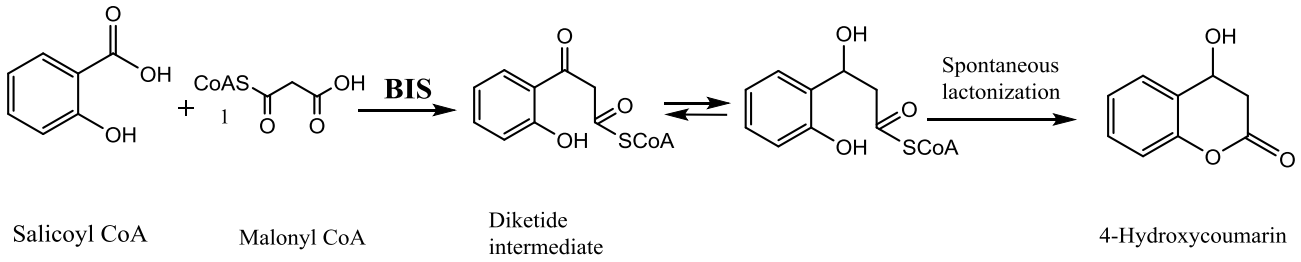


Fig I.17 The reaction catalyzed by biphenyl synthase upon use of *o*-hydroxybenzoyl-CoA as starter.

Benzoyl-CoA in general is a rarely used substrate among type III PKSs which is an intrinsic property of these enzymes. The second more challenging property is the use of the same substrate by two remarkable enzymes (BIS and BPS) to yield totally different groups of compounds. These observations made the reaction and the cyclization mechanisms very interesting to be involved in our research studies. Understanding the structures and comparing the active site cavities of both enzymes can explain the difference in their function. Interestingely, the residues which are responsible for the function diversity in type III PKSs are substituted in BIS and BPS. The crystallization of benzoic acid specific type III PKSs and the comparison with the crystal structures of CHS and STS will assist in deeper understanding the acceptance of small hydrophobic substrates like benzoyl-CoA by BIS and BPS and larger as coumaroyl-CoA by CHS and STS. The crystal structures of BIS and BPS disclose the differences in the shape and volume of the active sites pockets, which should indicate the reasons behind the differently utilized mechanisms and products formed, although using the same starter substrate.

2.3.2.4. Crystallization of other plant type III PKSs

A. 2-Pyrone synthase (2-PS)

It is similar to MsCHS2 by sharing 74% sequence identity at the amino acids level, although both enzymes differ in the starter substrate preference and the number of condensations utilized. The first crystallized pyrone synthase was cloned from *Gerbera hybrida* (daisy) (GhPS), which uses acetyl-CoA as starter substrate and produces a triketide intermediate by the addition of two extra acetyl units derived from malonyl decarboxylation. After cyclization, 6-methyl-4hydroxy-2-pyrone is formed, which presents the backbone for broader pyrone synthesis (Figure I.18).

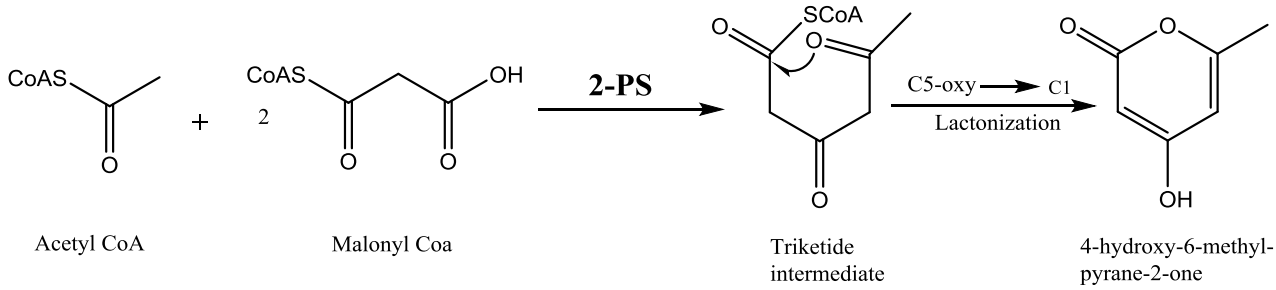


Fig I.18 The reaction catalyzed by 2-pyrone synthase.

The crystal structure of 2-PS in comparison to MsCHS2 demonstrated four different amino acids in the elongation pocket and they are Leu 202, Met 259, Leu 261, and Ile 343 in 2-PS and Thr 197, Ile 254, Gly 256, and Ser 338 in MsCHS2 respectively (Figure I.19). Such variations allowed the production of a triple mutant of MsCHS2, MSCHS2T197L/G256L/S338I which did not accept p-coumaroyl-CoA and resulted instead in methylpyrone as 2-PS. The size of the active site cavity of 2-PS was 274 Å in comparison to 923 Å of MsCHS2, which was explained by the larger size residues such as Leu and Ile, which allowed this enzyme to tolerate only acetyl-CoA as a substrate to form the corresponding tri-and not the tetraketide intermediate (Jez et al., 2000a). The small size cavity of 2-PS allowed it also to accept small CoA-thioesters like propionyl-CoA and benzoyl-CoA in vitro to produce 6-phenyl-4-hydroxy-2-pyrone, which composes the skeleton of some HIV-1 protease inhibitors (Austin & Noel, 2003).

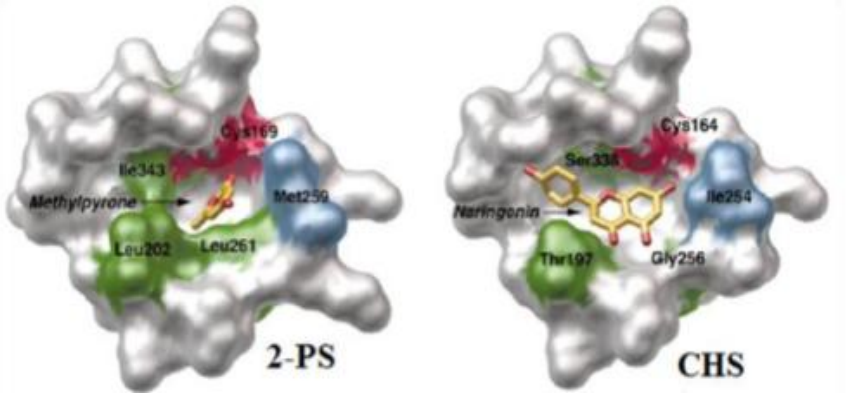


Fig I.19 The active site cavity of 2-pyrone synthase in comparison to the active site cavity of MsCHS2.

B. Benzalacetone synthase (BAS)

It was cloned from *Rheum palmatum* (RpBAS) and it shares 60-75% sequence identity at the amino acid level with other enzymes in the superfamily of CHS. It prefers 4-coumaroyl-CoA and condenses one molecule of malonyl-CoA to produce benzalacetone which provides the C6-C4 skeleton of phenylbutanoids (Figure I.20). These compounds are the active principle of medicinal plants having an anti-inflammatory effect such as gingerol in *Zingiber officinale* (Abe et al., 2001).

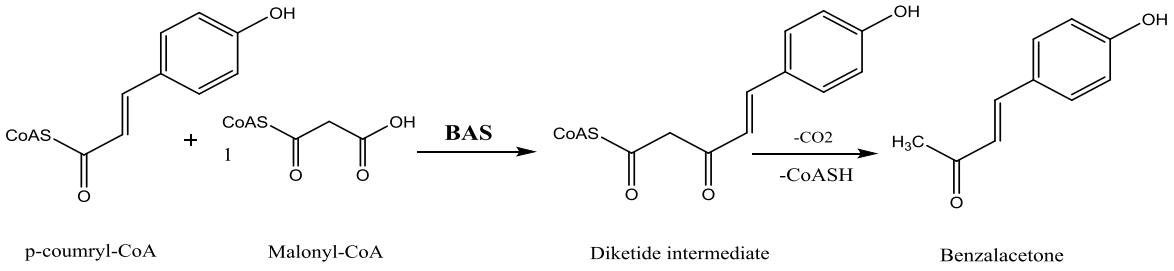


Fig I.20 The reaction catalyzed by benzalacetone synthase.

In comparison to the structure of MsCHS2, BAS lacks the active site residue Phe215. This observation promoted a mutation to replace Ile207 and Leu208 of BAS into the corresponding ones of MsCHS2, Leu214 and Phe215, respectively, which resulted in the conversion of BAS into CHS (Abe et al., 2003). Another mutant was constructed to replace Leu125, Ile207 and Leu208 of BAS to the corresponding ones of MsCHS2, Thr132, Leu214 and Phe215, respectively, to increase the chalcone formation but the result was a loss of the activity (Shimokawa et al., 2010). The replacement of Gly249 in BAS with Leu decreased the activity, which is the reverse of MsCHS2, whose G256L mutant did not accept p-coumaroyl-CoA at all. This hypothesized the presence of a second alternative coumaroyl-binding pocket (Abe et al., 2007b), which was proved by solving the crystal structure of RpBAS together with the chalcone-producing mutant Ile207Leu/Leu208Phe. This crystal structure also proved for the first time the nucleophilic function of the active site cysteine in polyketide intermediate binding, as the crystal structure of the monoketide coumarate intermediate bound to the wild type enzyme was solved (Figure I.21) (Morita et al., 2008, Morita et al., 2010a).

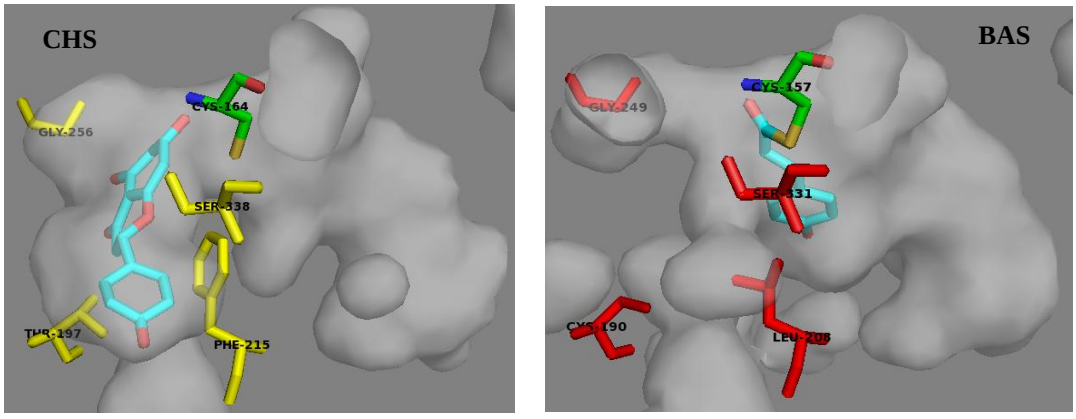


Fig I.21 The second coumaroyl-binding pocket of RpBAS. The different residues of RpBAS (red sticks) and MsCHS2 (yellow sticks) are highlighted. The monoketide intermediate and naringenin are shown as cyan sticks in RpBAS and MsCHS2, respectively.

C. **Curcuminoid synthase (CUS)**

It was cloned firstly from rice (*Oryza sativa*) (OsCUS) and shared 49% and 51% amino acid identities with MsCHS2 and RpBAS, respectively. It prefers 4-coumaroyl-CoA as a starter substrate and condenses only one molecule of malonyl-CoA to produce bisdemethoxycurcuminl. It also results in the formation of dicinnamoylmethane and curcumin after reactions using cinnamoyl-CoA and feruloyl-CoA, respectively (Figure I.22). Curcumin is a widely known curcuminoid because of its antiinflammatory and antitumor activities. It was found in the rhizome of *Curcuma longa*, which is used abundantly in Asia in the traditional medicine and also as food additives. The reaction mechanism is similar to BAS and involves the formation of β-keto acid but differs because of the use of another extender in addition to malonyl-CoA. The mechanism involves the decarboxylative condensation of the first moiety of malonyl-CoA with 4-coumaroyl-CoA to produce the diketide CoA intermediate, which undergoes further hydrolysis to produce the β-keto acid. This acid then functions as the second extender for another coumaroyl moiety. After another decarboxylative condensation reaction, bisdemethoxycurcumin is formed (Figure I.22) (Katsuyama et al., 2007).

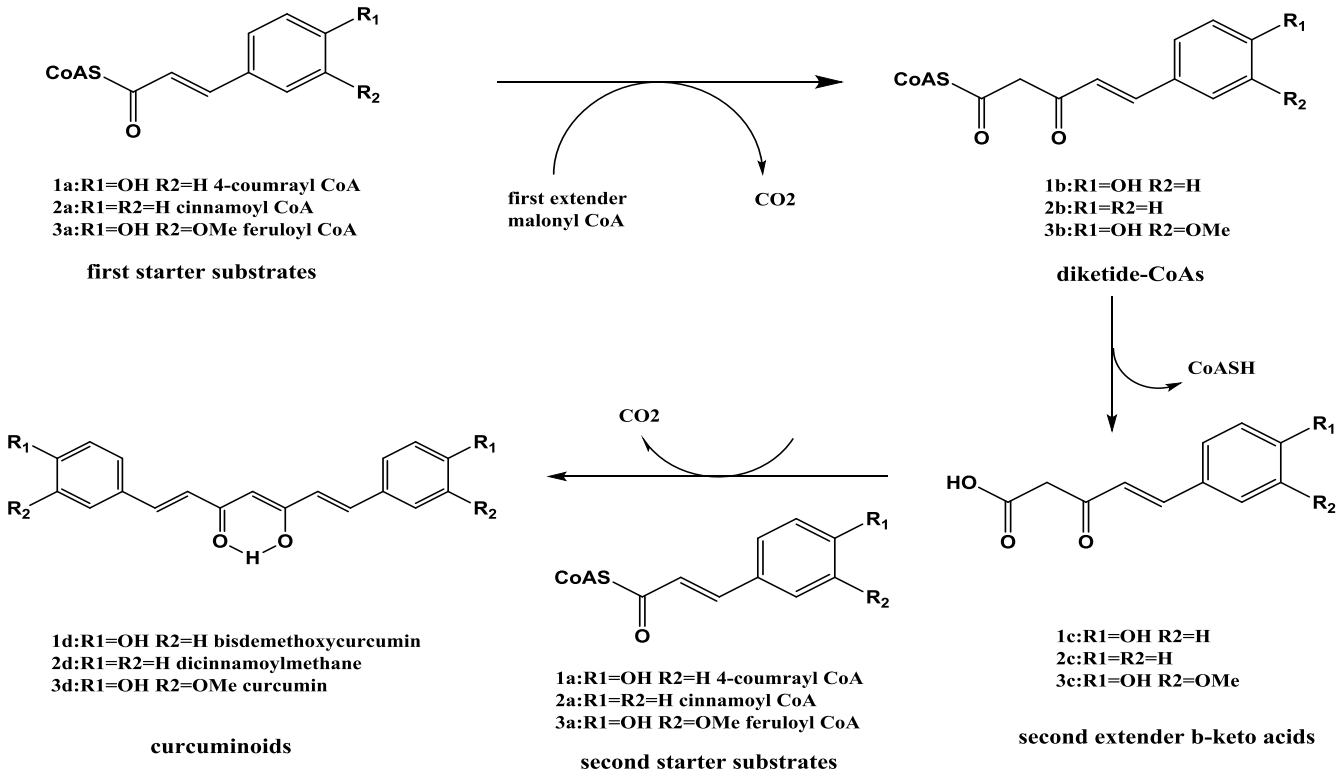


Fig I.22 The reactions catalyzed by curcuminoid synthase.

Interestingly, when the crystal structure of OsCUS was solved, it showed no coumaroyl-binding pocket as MsCHS2 and PsSTS. Instead, another expanded downward pocket was observed, which presumably tolerated the 4-coumaroyldiketide acid (Figure I.23) (Morita et al., 2010b).

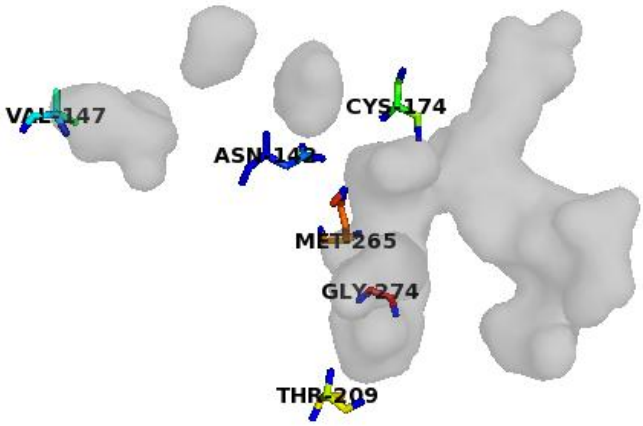


Fig I.23 The expanded downward pocket of OsCUS. This pocket accommodates the *β*-keto acid and has important residues that differ from MsCHS2.

A new curcumin synthase was cloned from turmeric *Curcuma longa* (ClCURS). It showed a unique acceptance of substrates lacking CoA, such as 3-oxooctanoyl-NAC. This suggested the involvement of a hydrophobic cavity surrounding Phe265 in the CoA binding tunnel. To prove this, a mutation was constructed on a residue surrounding the hydrophobic cavity Gly211.Glycine was replaced with the bulk residues Trp and Phe and both mutants showed weak activity in curcuminoids production. In addition, the crystal structure of the G211F mutant clarified the occupancy of the hydrophobic cavity by the phenyl group of Phe211 after mutation, which proves the importance of this cavity in β-keto acid binding and in curcuminoid synthesis (Figure I.24) (Katsuyama et al., 2011, Katsuyama et al., 2009).

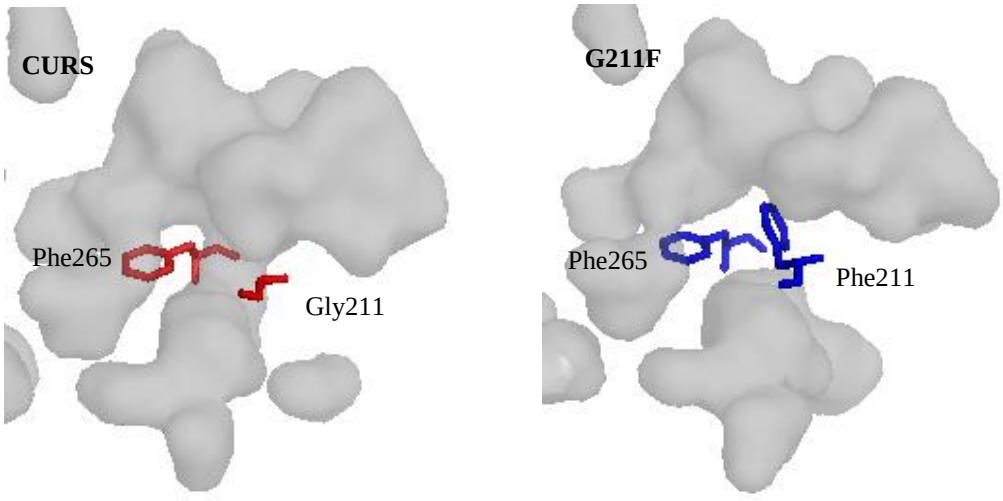


Fig I.24 The hydrophobic cavity of CURS. Gly (red) in CURS and Phe211 (blue) in its mutant CURSG211F are highlighted.

D. Acridone synthase (ACS)

It is involved in the synthesis of acridone alkaloids, which are known to be abundant in the Rutaceae family. Acridones have many activities including antiviral, antiplasmodial and antineoplastic properties, which are mainly related to their interference with DNA synthesis. ACS promotes the condensation of *N*-methylanthranilic acid and malonyl-CoA to produce 1,3-dihydroxy-N-methylacridone and this compound provides the skeleton for more complex acridone alkaloids (Figure.27) (Baumert et al., 1994, Junghanns et al., 1995). CHS and ACS are similar enzymes as they both catalyze the Claisen condensation mechanism using three molecules of malonyl-CoA to produce the corresponding tetraketide. ACS specifically forms a non-aromatic intermediate, which undergoes dehydration to lose a water molecule before aromatization (Figure I.25) (Springob et al., 2000).

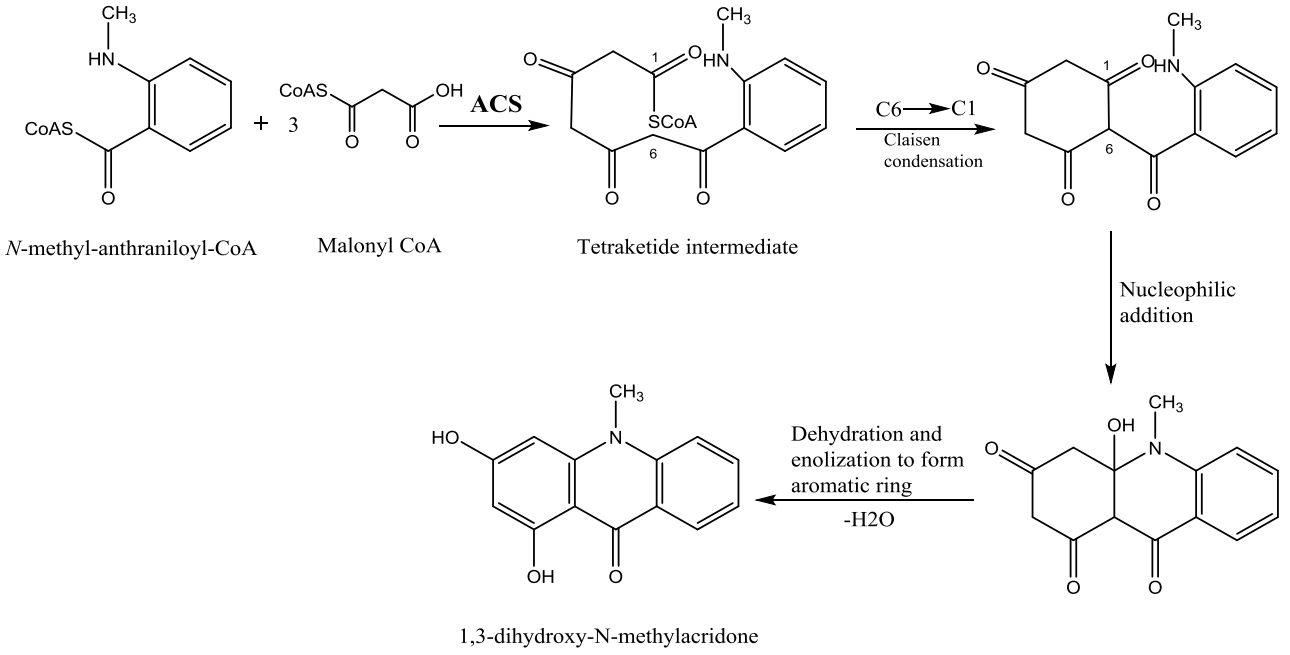


Fig I.25 The reactions catalyzed by acridone synthase.

The crystal structure of novel ACS from *Citrus microcarpa* (CmACS) has been solved recently(Mori et al., 2013). Structure analysis revealed the widening of the active site cavity entrance, which enhanced the acceptance of various substrates, explained by main residues differing from MsCHS2. For example, ACS has Val at position 265 instead of Phe and the loss of the aromatic ring changed the shape of the active site cavity. In addition, the side chain of Ser at position 132 instead of Thr protrudes outside the active site cavity and the side chain of Ala at position 133 instead of Ser is smaller and less bulky (Figure I.26).

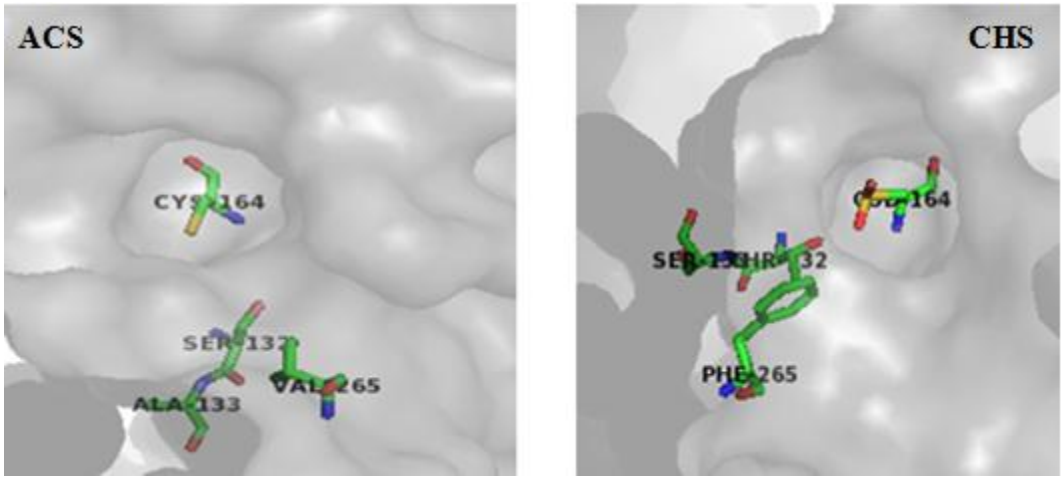


Fig I.26 The differences in the size of the active site entrances between CmACS and MsCHS2.

Importance sites in CmACS (Ser132, Thr194 and Thr197) were mutated to Met, Met, and Tyr, respectively, which are the corresponding ones in *C. microcarpa* quinolone synthase (CmQNS). All single mutants produced converted ACS into functional QNS, as they accepted only *N*-methylanthraniloyl-CoA to produce quinolone after one condensation reaction with malonyl-CoA. All the mutants did not accept p-coumaroyl-CoA as the wild type enzyme did (Figure I.27) (Mori et al., 2013).

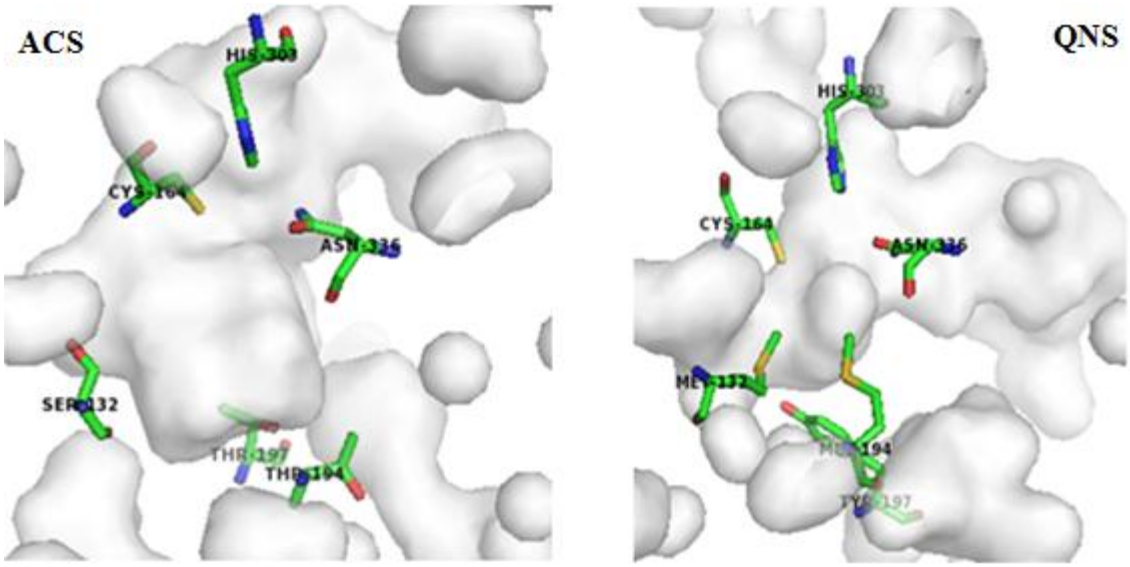


Fig I.27 The important residues that differ between *Citrus microcarpa* ACS and QNS.

E. Quinolone synthase (QNS)

It prefers *N*-methylanthraniloyl-CoA as starter substrate and catalyzes one condensation with malonyl-CoA to produce 4-hydroxy-1-methyl-2-quinolone (Figure I.28). Quinolone alkaloid derivatives are distributed in plants of Rubiaceae, such as antimalarial quinine and skimmianine that has cytotoxic and antiplatelet aggregation effects (Resmi et al., 2013).

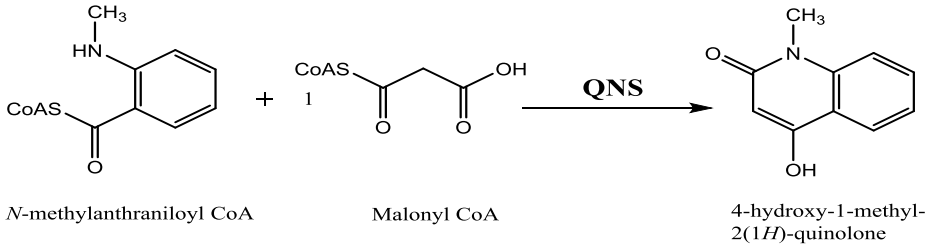


Fig I.28 The reaction catalyzed by quinolone synthase.

RpBAS was the first enzyme to produce quinolone by its acceptance of N-methylanthraniloyl-CoA (Abe et al., 2006a). Recently, quinolone synthase was first cloned from the root of *Aegle marmelos* (AmQNS), which is known for its abundant quinolone alkaloids. Mutational experiments proved the evolution of QNS from CHS. The substitution of two conserved residues in AmQNS (Ser132, Ala133) into the corresponding ones of MsCHS2 (Thr132, Ser133) led to the conversion of QNS into CHS, as the mutant resulted in naringenine formation after reaction with p-coumaroyl-CoA (Resmi et al., 2013). In 2013, the structure of CmQNS was solved and it revealed the structure differences as the basis for functional variation of this enzyme. Especially, it has different bulk residues (Met132, Ala133, Leu265) instead of Ser132, Ala133, Val265, respectively, in comparison to the previously cloned AmQNS. This steric hindrance changed the shape and the size of the active site cavity and allowed the Leu267 to permit behind Leu265 to participate in the entrance of the active site. In comparison to MsCHS2, the absence of the phenyl ring of Phe265 allowed the other Phe215 to protrude towered Leu265 and to widen the active site entrance to reach 47 Å, which is 2.8 times larger than that of MsCHS2 (17 Å) (Figure I.29).

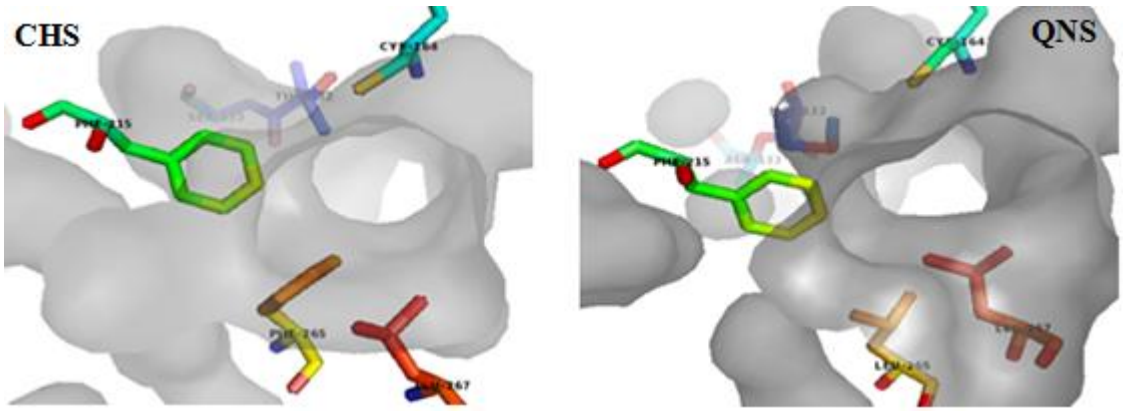


Fig I.29 The variation in the substrate entrances of MsCHS2 and CmQNS. The orientation of the different residues affects the shape and size of the entrances in both enzymes.

Two bulk Met residues in CmQNS replace Thr132 and Thr194 of MsCHS2, which leads to the absence of the coumaroyl binding pocket. The presence of Tyr instead of Thr197 and the presence of Ala instead of Gly256 decreased the size of the active site cavity to 290Å, which is 2.6 times smaller than that of MsCHS2. Mutational experiments proved the critical role of Tyr197 in such a change, as it converted QNS into BAS after replacing it with Ala to allow the mutant to also accept 4-coumaroyl-CoA as starter substrate (Figure I.30) (Mori et al., 2013).

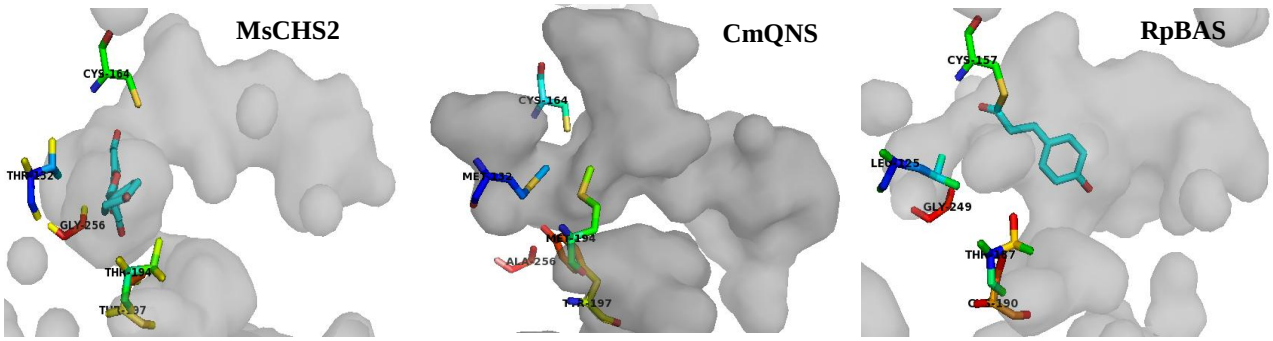


Fig I.30 The different active site geometries of MsCHS2, CmQNS and RpBAS. The important different residues surrounding the active site cavity among all enzymes are shown as sticks. Monoketide intermediate and naringenin are shown as cyan sticks in RpBAS and MsCHS2, respectively.

F. Pentaketide chromone synthase (PCS)

This plant PKS condenses five molecules of malonyl-CoA to produce 5,7-dihydroxy-2-methylchromone. This aromatic pentaketide is the main nucleus in the synthesis of the furochromones kehellin and visnagin, which are known of their anti-asthmatic effects (Figure I.31). It was cloned from *Aloe arborescens* (AaPCS) and the sequence of its amino acids has about 58% identity with MsCHS2. Sequence comparison showed the substitution of three important residues of MsCHS2 (Thr197, Gly256 and Ser338) with Met207, Leu266 and Val351 of AaPCS, respectively.

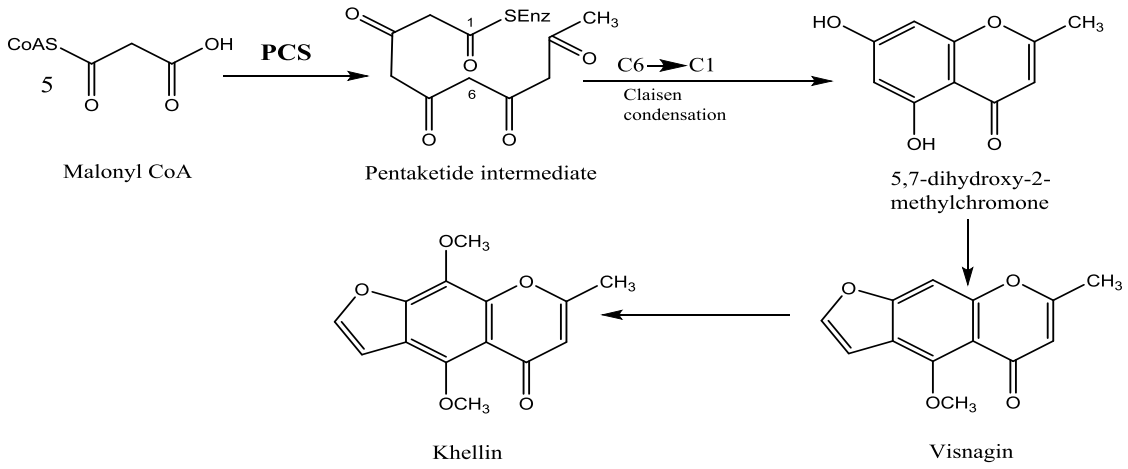


Fig I.31 The reactions catalyzed by pentaketide chromone synthase and subsequent enzymes.

Similar residues were found in the amino acids sequence of octaketide synthase from *Aloe arborescens* (AaOKS), except for a Gly at position 207 instead of Met. Based on this a mutation was done on PCS and the produced single site mutant M207G was able to condense eight molecules of malonyl-CoA and transformed PCS into functionally active OKS (Abe et al., 2005b). After solving the crystal structures of AaPCS and its mutant M207G, a new hidden pocket was found behind the active site cavity of M207G in comparison to PCS, which explains the longer chain production. A triple mutant was further constructed by substitution of extra two aromatic bulk residues (Phe80 and Tyr82) at the bottom of the new pocket into Ala. The produced triple mutant (F80A/Y82A/M207G) produced unnatural novel nonaketide naphthopyrone after condensation of nine molecules of malonyl-CoA. The volume of the active site cavity of the triple mutant is four times larger than the one of the wild type PCS (Figure I.32) (Morita et al., 2007).

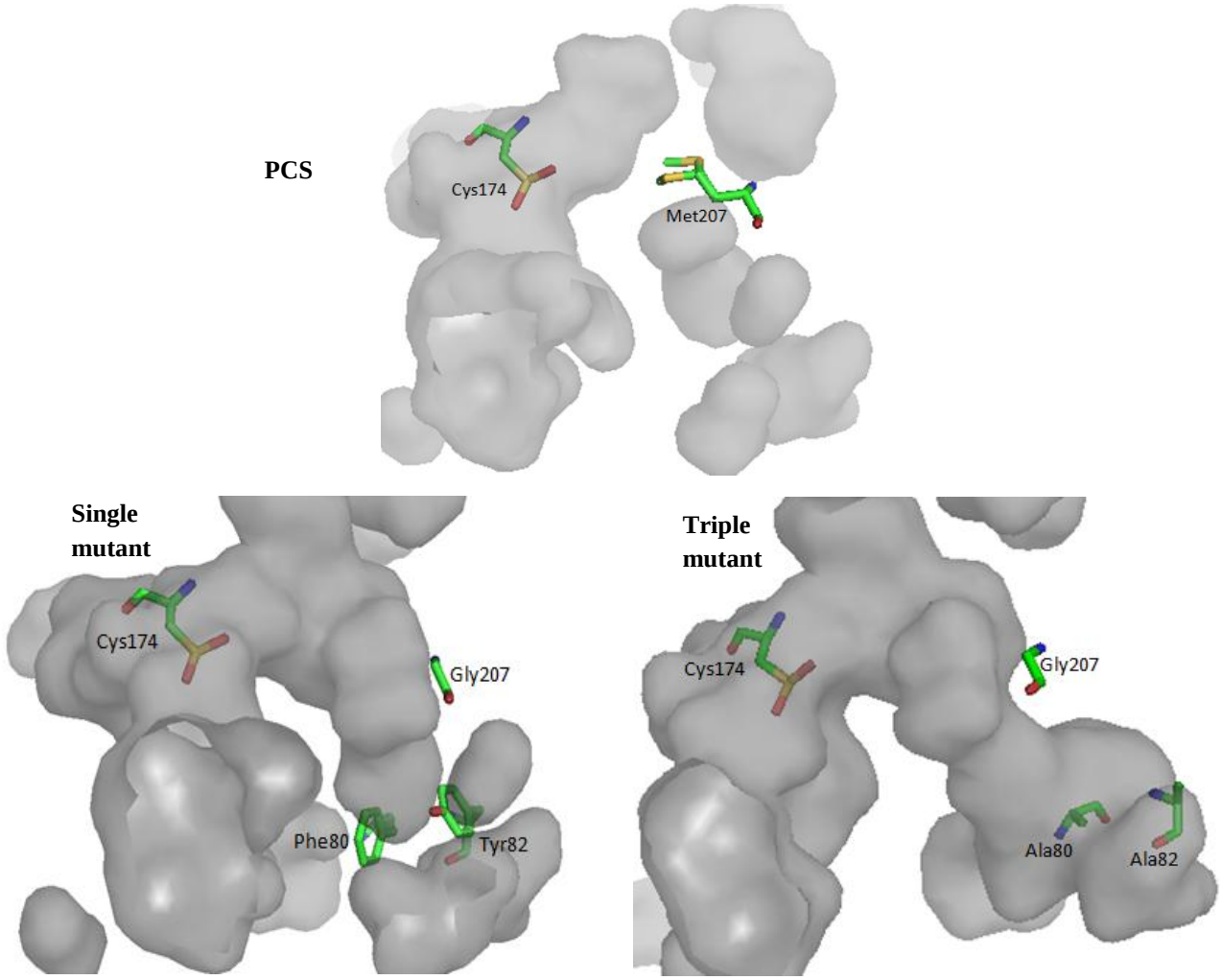


Fig I.32 The differences in the size of the active site pockets of PCS and its mutants. AaPCS and its single mutant M207G had smaller active site pockets in comparison to the expanded large pocket of its triple mutant (F80A/Y82A/M207G). The crystal structure of the triple mutant was not available in PDB, but was constructed using PyMOL.

G. **Octaketide synthase (OKS)**

It was cloned from *Aloe arborescens* (AaOKS) and it condenses eight molecules of malonyl-CoA resulting in the formation of the octaketides SEK4 and SEK4b (Figure I.33). The *Aloe* plant is known of its abundance of important longer chain medicinal aromatic polyketides, such as aloenin (hexaketide), aloesin (heptaketide) and barbaloin (octaketide), but it does not produce SEK4 or SEK4b. So it was suggested that OKS is somehow involved in vivo in the biosynthesis of anthrones.

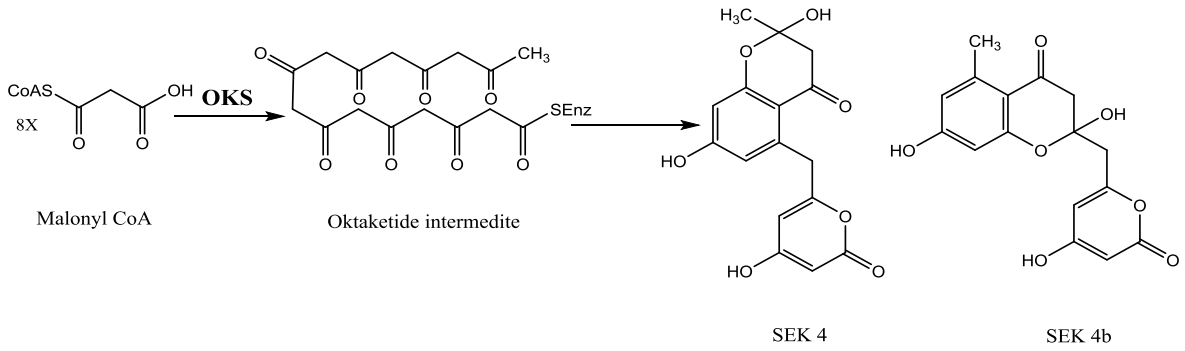


Fig I.33 The reaction catalyzed by octaketide synthase.

At the amino acids level, it shares about 91% identity with AaPCS and 60% identity with MsCHS2. As in PCS, three conserved residues (Thr197, Gly256, and Ser338) were substituted with Gly, Leu, and Val, respectively, in OKS. Gly207 was subjected to mutation and the resulting mutant G207M formed unnatural pentaketide 2, 7-dihydroxy-5-methylchromone. So, it converted OKS into functional PCS (Abe et al., 2005a). Another mutant N222G could condense ten molecules of malonyl-CoA to produce the unnatural compound C20 decaketide benzophenone SEK15. The solved structure of N222G showed an increase in the cavity volume from 652 Å to 693 Å which correlated to the produced longer polyketide. The residue Phe66 in N222G was further mutated with smaller residues, only the double mutant F66L/N222G was more interesting because another unnatural longer polyketides was produced namely dodecaketide naphthophenone TW95a after condensation of twelve molecules of malonyl-CoA. The volume of the cavity of F66L/N222G increased to 748 Å which explained the production of the C24 dodecaketide chain (Figure I.34)

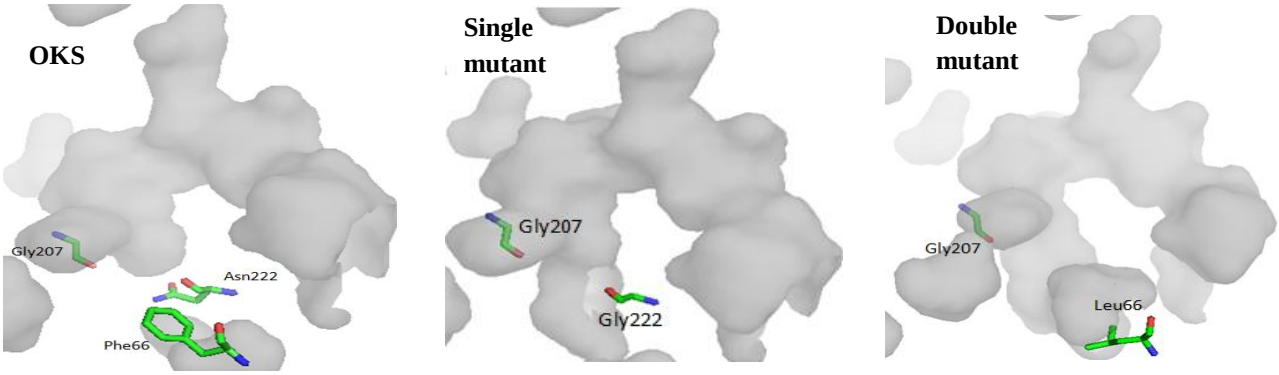


Fig I.34 The different active sites pocket sizes of OKS and its mutants. The double mutant F66L/N222G has a larger expanded pocket in comparison to N222G and OKS. The crystal structures of the single and double mutants were not available in PDB, but were constructed using the PyMOL program.

The dodecaketide naphthophenone TW95a proves the involvement of the CHS superfamily in the synthesis of anthrones in the *Aloe* plant. This C24 dodekaketide is the longest known polyketide produced by plant type III PKSs (Figure I.35) (Wanibuchi et al., 2011)

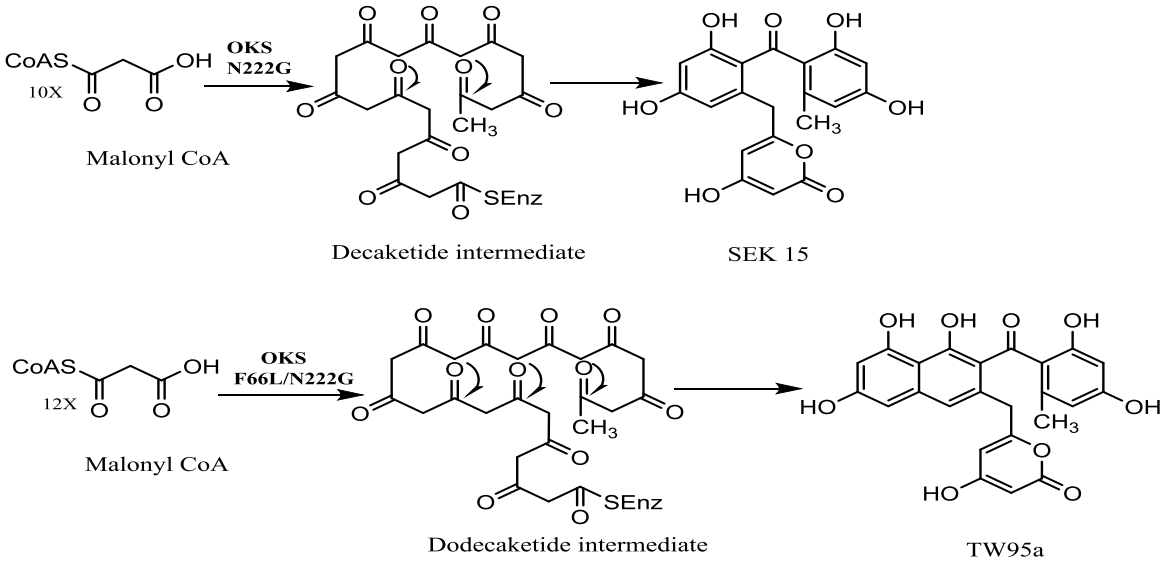


Fig I.35 The formation of the longest known plant polyketide TW95a.

2.3.3. Bacterial and fungal type III PKSs

Type III PKSs were thought to be plant-specific until 1999 when a new gene called *rpp*A was cloned from the soilliving filamentous bacterium *Streptomyces griseus*. This gene encoded a protein that shows significant similarity to CHS. RppA protein condenses five molecules of malonyl-CoA resulting in a pentaketide intermediate, which cyclizes to 1,3,6,8-tetrahydroxynaphthalene (THN) (Figure I.36).

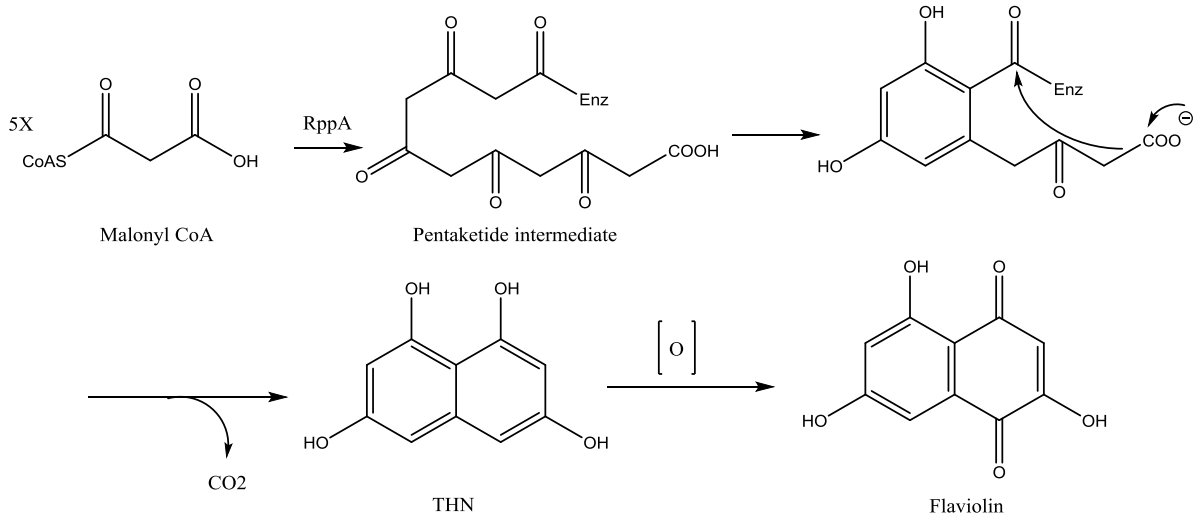


Fig I.36 The reaction catalyzed by RppA. The enzyme produces THN and the instability of this product results in the formation of flaviolin due to oxidation.

Unlike CHS, RppA did not provide alternative products after the co-incubation of malonyl-CoA with either p-coumaroyl-CoA or acetyl-CoA. This indicated that, RppA is a CHS-related PKS which utilizes only malonyl-CoA as the starter unit (Funa et al., 1999). Since 1999, many other related genes were found in various bacteria and revealed that type III PKSs are distributed in a variety of bacteria (Moore & Hopke, 2001, Gross et al., 2006). An important example is the new family of type III PKSs that was found in *Mycobacterium tuberculosis*, which enfolds mainly PKS18 and PKS11. Both enzymes are involved in the biosynthesis of long-chain α-pyrones as they showed higher specificity toward long-chain aliphatic acyl-coenzyme A (C12 to C20) substrates. This is unprecedented in the CHS superfamily, as octanoyl-CoA considered to be the largest starter unit used by CHS and THN (Saxena et al., 2003). In general, the acceptance of long-chain acyl-CoA starters by bacterial type III PKSs is strongly correlated to their large active sites cavities. For example, the crystal structures of 1,3,6,8-tetrahydroxynaphthalene synthase (THNS) demonstrated the prescence of a novel cavity expanded into the floor of the main pocket (PEG-binding tunnel). Moreover, THNS, as all known bacterial type III PKSs, features three additional active site cysteines, besides the catalytic Cys at the positions 106, 168, and 171 (corresponding to 132,194 and197 of MsCHS2, respectively). These positions are mostly occupied by threonine residues in MsCHS2 and other plant type III PKSs (Figure I.37). They are less bulk and ideal for the ability of THNS to utilize small starters while providing adequate volume for the additional elongation steps of the polyketide chain. Mutagenic analysis of THNS proved the unexpected vital function of Cys106. It showed an important role of utilizing THNS dual cyclizations reactions (adol and Claisen) for THN biosynthesis (Austin et al., 2004b). The crystal structure of PKS11 also showed a similar expanded hydrophobic tunnel close to the active sites Cys, which accommodates the large reaction substrates (Gokulan et al., 2013).

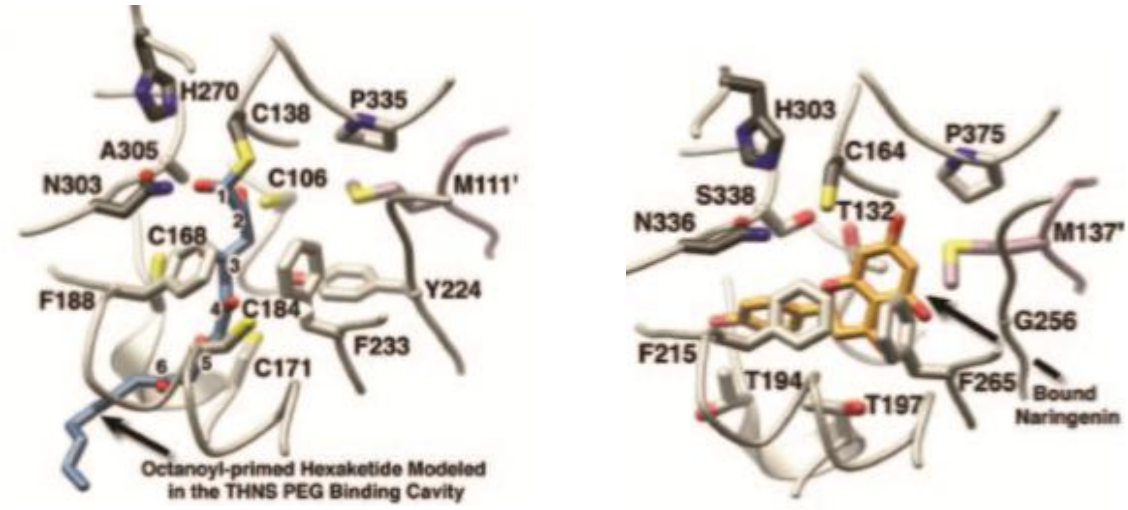


Fig I.37 The active site cavity of THNS. On the left, the active site cysteines are presented as sticks. They are bulk residues and oriented to accommodate the modeled hexaketide in the PEG-binding tunnel. On the right, the pocket of MsCHS2 accommodating naringenin is wider in comparison to THNS.

Single amino acids substitutions could reverse the cyclization specificity of two bacterial type III PKSs (ArsB and ArsC) obtained from the nitrogen-fixing soil bacterium, *Azotobacter vinelandii*. The ArsB W281G synthesized alkylpyrone but not alkylresorcinol. In contrast, the ArsC G284W synthesized alkylresorcinol with a small amount of alkylpyrone. The crystal structures of ArsC and its mutant ArsC G284W observed the steric wall of the mutant that reduced the volume of its active site cavity (Figure I.38). This hypothized that the polyketomethylene intermediate can be folded to a suitable form for aldol condensation only in such a small cavity of ArsCG284W. This folding is not suitable for lactonization as in the wild-type enzyme ArsC (Satou et al., 2013).

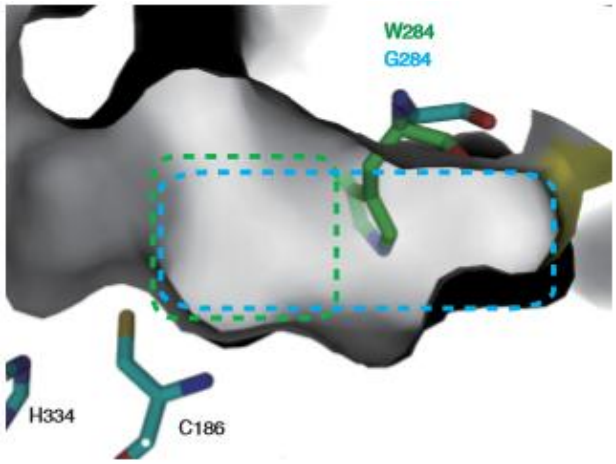


Fig I.38 The active site cavities of ArsC and its mutant ArsC G284W. The pocket size of the mutant is smaller because of the prescence of W284 instead of G284. This helps in the perfect folding of the polyketide intermediate and allows the utilization of an aldol rather than a lactonization mechanim by ArsC G284W.

Another intersting bacterial type III PKS, germicidin synthase (Gcs) from *Streptomyces coelicolor*, demonstrated broad substrate acceptance for acyl groups linked through a thioester bond to either coenzyme A (CoA) or acyl carrier protein (ACP). The catalytic efficiency for acyl-ACP was even 10-fold higher than that for acyl-CoA. In contrast, plant type III PKSs tend to utilize a larger variety of acyl-CoA and not acyl-ACP, which is a general feature of type I and II PKSs (Chemler et al., 2012).

Not only bacterial type III PKSs were characterized, but in 2007, the first fungal type III PKS from *Neurospora crassa* was determined. This novel type III PKS catalyzes the synthesis of pentaketide alkylresorcylic acid from a long-chain fatty acid CoA ester and four molecules of malonyl-CoA. It prefers C_{18} stearoyl-CoA as starter and is called 2'-oxoalkylresorcylic acid synthase (ORAS) (Funa et al., 2007). It is the first type III PKSs that catalyses the formation of pentaketide resorcylic acid as most of other type III PKSs produce tetraketide resorcinols. Since then other many fungal type III PKSs were identified. For example *Aspergillus niger* AnPKS that showed specific acceptance of acyl-CoAs with branched chains such as isobutyryl-CoA and isovaleryl-CoA. It did not accept crotonyl-CoA but accepted methylcrotonyl-CoA. It accepted also aromatic CoAs such as benzoyl-CoA and phenylacetyl-CoA.

Generally, fungal type III PKSs produced triketide, tetraketide and even petaketide pyrones. They also produced tetraketide, pentaketide and hexaketide resorcinol/resorcylic acis. Bacterial types III PKSs produced more specific products, such as naphthalene by THNS and phloroglucinol by PhID (from *Pseudomonas florescens*). Fungal type III PKSs utilize malonyl-CoA as extender except for CsyB from *Aspergillus oryzae*, which uses in addition acetoacetyl-CoA. Bacterial Sco7671 from *Streptomyces coelicolor* accepted methylated and ethylated malonyl-CoA as extenders (Hashimoto et al., 2014).

Beyond bacteria and fungi, several species of brown algae display genes that are homologs of type III PKSs. Recently in 2013, the crystal structure of the Esi-PKS1 protein of the brown alga *Ectocarpus siliculosus* was solved. Esi-PKS1 functions as a phloroglucinol synthase using malonyl-CoA as a substrate. It showed acceptance also of longer chain starters like lauroyl-CoA. The structure demonstrated a specific tunnel for binding of long-chain acyl-CoAs. In addition, two GF (G164F and G190F) are located in two loops close to the surface of the enzyme. The positions of these residues allowed the formation of a new pocket at the dimer interface. This pocket is absent in plant type III PKSs as it is occupied mostly by an aromatic residue (Tyr160 of MsCHS2) instead of the small residue G190 of Esi-PKS1 (Meslet-Cladiere et al., 2013).

It is important in the future to combine all the information about the structures and reaction mechanisms of type III PKSs from all known sources (plants, bacteria, fungi and even algae) to utilize the unique function of this group of enzymes for the production of desired polyketides.

2.3.4. Reaction mechanism of type III PKS

The reaction mechanism of MsCHS2 includes different functions, which resemble the action of many enzymes. It starts with a transferase for loading the starter p-coumaroyl-CoA onto the SH group of the catalytic Cys164, followed by a decarboxylase for activating the extender malonyl-CoA. Finally, cyclase and aromatase, which, after addition of C_2 units and formation of C-C bonds in the linked CoA chain, catalyze intrinsic intramolecular Claisen-type condensation of the linear tetraketide and form the final product. Actually CHS and all type III PKSs only need three essential residues (Cys164, His303 and Asn336) (numbering from MsCHS2) to perform the mentioned functions. The basic N-2 of His removes a proton from the SH group of Cys, which in turn activates the thiolate anion and provides a nucleophillic attack to the carbonyl group of the starter substrae to release the CoA moiety. The protonated nitrogens of His and Asn stabilize the carbonyl oxygen of malonyl-CoA to facilitate the decarboxylation reaction and production of acetyl carbanion, which attacks the carbonyl group of the starter-Cys complex and a diketide forms. This growing diketide will be taken again by the Cys-thiolate to allow a second molecule of malonyl-CoA to go in and further decarboxylation will occur to allow the addition of a new acetyl moity to the diketide.

Repetition of this process will occur untill the production of the tetraketide intermediate. Still bound to the enzyme, this intermediate undergoes cyclization and aromatization to give the final product (Figure I.39) (Abe, 2008, Abe & Morita, 2010).

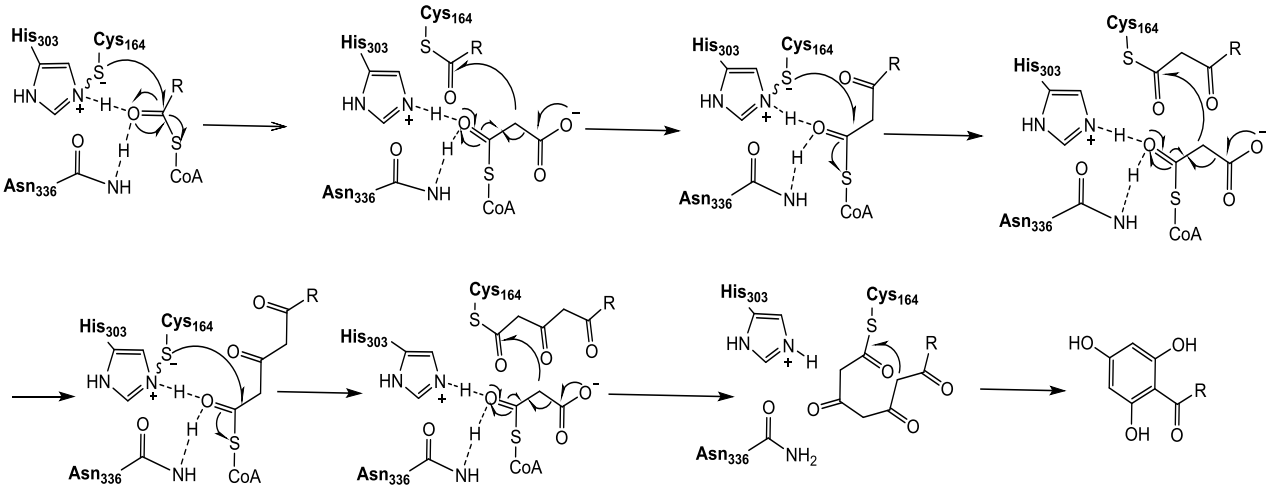


Fig I.39 The reaction mechanism of type III polyketide synthases.

2.3.5. Mechanism of Claisen condensation and aldol-type cyclization

The crystal structures of MsCHS2 and PsSTS demonstrated the significance of an important residue in the Claisen-aldol mechanisms variation (Thr132 in MsCHS2 and Thr135 in PsSTS). It is located close to the active site cavity and after mutation it showed an important effect on the production of resveratrol. In MsCHS2, the hydroxyl group of this residue forms hydrogen bonding with glutamate residue Glu192 at the back of the active site cavity but in PsSTS the changed conformation of the residue positions this hydroxyl group to form new hydrogen bonding by stabilizing a water molecule with Ser 338 (Figure I.40). This bonding pattern is similar to the catalytic activity of type II thioesterase discovered in *E. coli*. To prove this activity, an extra mutation was carried out on a 18xCHS mutant mentioned in (1.2.3.2) and each mutant of 18xCHS (+1) (Thr132 into Ala and Glu192 into Gln) showed an increase in the production of chalcone rather than stilbene (Austin et al., 2004a).

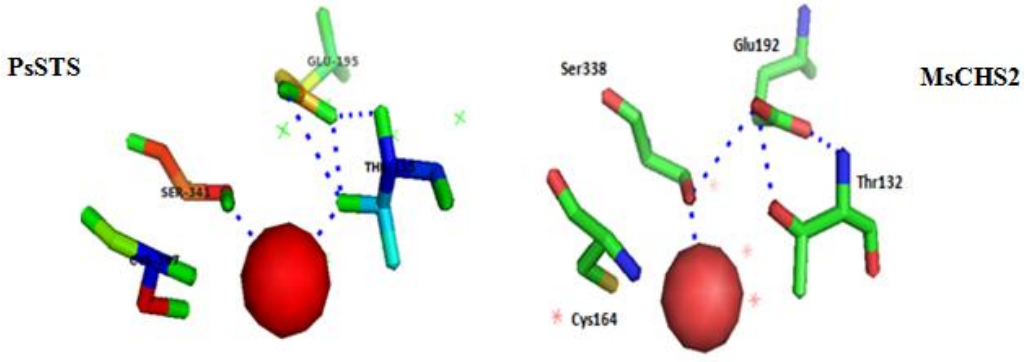


Fig I.40 The catalytic water molecule (red sphere) in PsSTS and MsCHS2. The important residues responsible for the catalytic water stabilization by hydrogen bond network formation (blue dotted lines) are displayed.

RpBAS and OsCUS utilize an aldol similar mechanism, which involves the decarboxylation of β-keto acids. Thr135 was not found in RpBAS and OsCUS and replaced by Leu and Asn respectively.The formation of the β-keto acid was correlated to the presence of a water molecule in the crystal structures of both enzymes but in different location compared to the one of PsSTS. In RpBAS this water forms a hydrogen bond network with the catalytic residues Cys157, His296 and Asn329 (Figure I.41). This water molecule was suggested to perform the thioesterase function, by which it cleaves the thioester bond of the diketide intermediate that will undergo further decarboxylation to produce benzalacetone (Morita et al., 2008, Morita et al., 2010a).

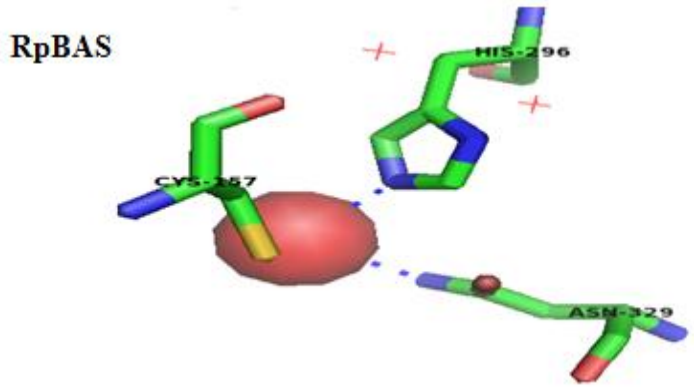


Fig I.41 The catalytic water molecule (red sphere) in RpBAS. The catalytic triad residues responsible for its stabilization by hydrogen bond network formation (blue dotted lines) are displayed.

In the crystal structure of OsCUS, another putative water molecule forms a hydrogen bond network with Asn142 (Thr135 in PsSTS), Ser351, Tyr207 and Glu202 in the active site cavity and its function was correlated to the hydrolysis and the release of the diketide intermediate as 4-coumaroyldiketide acid (Figure I.42) (Morita et al., 2010b).

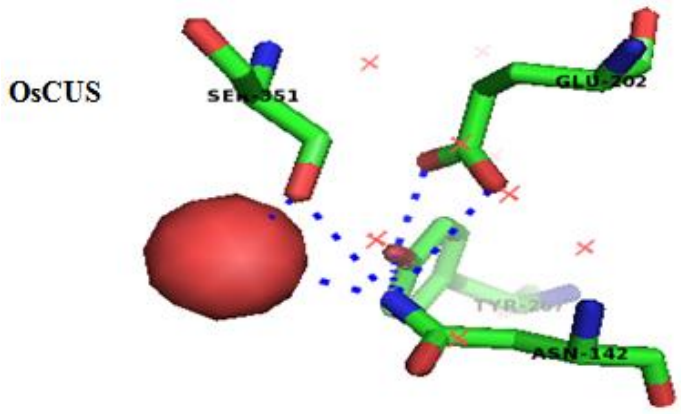


Fig I.42 The catalytic water molecule (red sphere) in OsCUS. The important residues responsible for its stabilization by hydrogen bond network formation (blue dotted lines) are displayed.

3. Protein crystallization and the impact of His_6-tag

Protein crystallization requires the creation of a supersaturation state. Supersaturation (non equilibirium state) occurs when the protein exceeds its solubility limit in the solution, which forces the solution to reequilibrate by the formation of a solid state (nucleus of a crystal). A stable nucleus in the supersaturation solution will grow until equilibrium is regained. The force that enhances the supersaturation or the growth of the crystal is the modification of the protein or the undersaturated solution poperties, i.e. high protein concentration, pH change, the addition of ligands and the use of precipitants like salts and polymers (Figure I.43) (McPherson & Gavira, 2014). More simply, crystallization is the choice among many probabilities that allow the conditions to proceed from equilibrium zone into supersaturation zone.

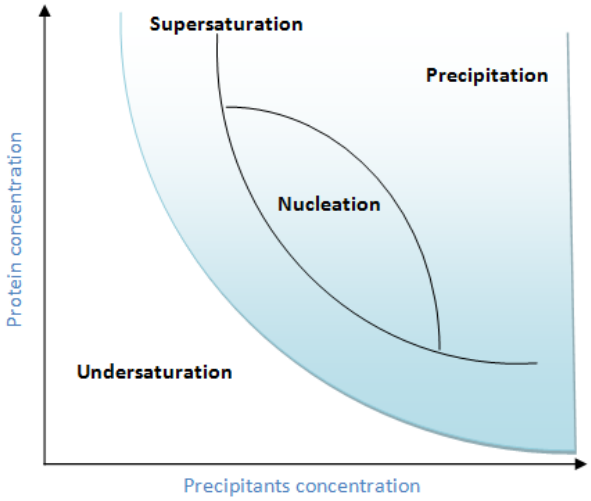


Fig I.43 The phase diagram for protein crystallization. It shows the different paths of a crystallization screen. Protein crystals can only grow from a supersaturated solution.

The His_6-tag facilitates the binding of the expressed target protein to a nickel-affinity column. It was commonly thought that these tags have no influence on the structure or the function of the protein. However, they may either affect the solubility of the purified protein or increase its aggregation. Therefore some crystallographers prefer to cleave the tags before crystallization screens because they may hinder protein crystallization. One study compared the crystal structures of some proteins with and without His_6-tags and concluded that the tags had no significant effect on the structure of the native protein. Actually, for some kinds of proteins it was necessary to crystallize them in the presence of the tag. In the future, it may be possible to design specific crystallization tags that may enhance protein crystallization and structure determination (Carson et al., 2007). Proteins behave differently and it is worth trying both constructs, i.e. either with tag or without tag, because it is never known which one will crystallize.

4. Aims of this study

- Optimization of expression and purification for high yields of His_6-tagged proteins:
  - Biphenyl synthase from *Sorbus aucuparia* (SaBIS1) and *Malus domestica* (MdBIS3).
  - Benzophenone synthases from *Hypericum androsaemum* (HaBPS) and *Hypericum sampsonii* (HsBPS).
  - Single mutants HaBPST135L and HsBPST135K.
- His_6-tag cleavage from purified His_6-tagged proteins using TEV protease.
- Crystallization of the purified enzymes and co-crystallization with substrates (CoA, benzoyl-CoA) and products (3,5-dihydroxybiphenyl, 2, 4, 6- trihydroxybenzophenone) in cooperation with Prof. Wulf Blankenfeldt and Dr. Peer Lukat, HZI, Braunschweig.
- Elucidation of the structural basis and explanation of the two different cyclization mechanisms used by BIS and BPS.
- Disclosing the structural basis of HaBPS and HsBPS.
- Finding structural explanations for the mutants HaT135L and HsT135K.

II. Materials

1. Chemicals

Chemicals, unless otherwise mentioned were purchased from the following companies: Roth, Sigma-Aldrich, Applichem, Fischer Scientific and Fluka. Deionized water was supplied from a Milli-Q water purification system (Sartorius, Germany) and used in the preparation of all aqueous solutions. All solutions were autoclaved by 120°C for 20 min. Solutions of thermolabile compounds as ampicillin and IPTG were sterile filtered and added to autoclaved solutions under sterile conditions. All salts required for the bacterial culture media were supplied from Roth or Applichem.

Chemicals	**Supplier**
For bacterial culture media	
Yeast	Applichem
Pepton(casein)	Roth
NaCL	Roth
Agar	Applichem
Glycerol	Roth
Stationary phases used in protein purification	
Sephadex G-25 in PD10 desalting columns	GE Healthcare
Nickel-nitrilotriacetic acid Agarose resin	Qiagen
Sephacryl S300 HR	GE Healthcare
Superdex S200 10/300 GL	GE Healthcare
Mono-Q HR 5/5 anion exchange column	GE Healthcare
Substrates used for enzyme assay	
Benzoyl-CoA and malonyl-CoA	Sigma-Aldrich
Antibiotics	
Ampicillin	Roth
Chloramphenicol	Applichem

Chemicals	**Supplier**
Reagents for biochemistry and molecular biology	
IPTG	Applichem
Immidazole	Roth
Sodium dihydrogen phosphate	Roth
Potassium dihydrogen phosphate	Roth
Sodium hydroxide	Roth
Potassium hydroxide	Roth
Acetic acid	Roth
Hydrochloric acid 37%	Riedel-de-Haen
o-Phosphoric acid	Roth
Tris-HCl	Roth
DNTPs	Thermo Scientific
Organic solvents	
Methanol	Fischer-Scientific
Ethanol absolute	Sigma Aldrich
Ethylacetate	Fischer-Scientific
Chloroform	Fischer-Scientific
Isopropanol	Acros Organics
Reagents for DNA and Protein gel electrophoresis	
peqGold universal agarose	Peqlab
Ethidium bromide	Roth
Medori green	NIPPON Genetics EUROPE
Acrylamide/Bisacrylamide 30%	Bio-Rad
TEMED	Bio-Rad
Glycin	Roth
SDS	Roth
β-mercaptoethanol	Roth
Bromophenol blue	Sigma-Aldrich
Commassie-blue R250 and G250	Merck
Ladders	
Gene Ruler DNA ladder Mix	Thermo Scientific
Page Ruler Unstained Protein Ladder	Thermo Scientific

2. Nutrient media, solutions and buffers

2.1. Nutrient media for bacterial culture

Name	Composition	
SOC Media	Peptone (Casein)	20 g/l
	Yeast extract	5 g/l
	1 M NaCl	10 ml/l
	1 M KCl	10 ml/l
	Autoclave and addition of the following sterile filtered solutions	
	2 M Mg^2	10 ml/l
	2 M Glucose	10 ml/l
LB Media	Peptone (Casein)	10 g/l
	Yeast extract	5 g/l
	NaCl	10 g/l
For solid media	Agar	1% (w/v)
For induction of protein expression	IPTG (Isopropyl-β-D-thiogalactopydranoside)	0.12 g/ml (0.5 M)
	Final concentration in bacterial culture	(0.1-0.5 mM)
Antibiotic for selection of transformed bacteria	Ampicillin (0.269 M)	100 mg /ml aqueous soln.
	Chloramphenicol (0.093 M)	30 mg/ml soln. in ethanol
For preservation of bacterial culture	Glycerol:LB medium (60:40)	250 µl
	Bacterial culture	750 µl

2.2. Buffers and solutions

2.2.1. Buffers for DNA- gel electrophoresis

50X TAE buffer	2 M Tris –HCl 0.05 M EDTA pH 8, adjusted by glacial acetic acid

1% agarose gel was prepared to examine DNA products.

2.2.2. Buffer for enzyme assay

Buffer KH2PO4	100 mM	1.36 g
	Water	ad 100 ml
	pH 7, adjusted by KOH	

2.2.3. Buffers for plasmid isolation (miniprep)

Buffer I (pH 8)	Tris-HCl	50 mM	1.5 g/250 ml
	EDTA	10 mM	0.93 g/250 ml
	RNase A	10 µl/ml	
	RNase A was added freshly before use		
Buffer II	NaOH	0.2 M	2 g/250 ml
	SDS	1% (w/v)	2.5 g/250 ml
Buffer III (pH 5.5)	Potassium acetate	2.55 M	62.57 g/250 ml
	Glacial acetic acid to adjust pH		

2.2.4. Solutions and buffers for protein gel electrophoresis (SDS-PAGE)

Staking gel (for 2 small gels)	Water	2.72 ml
	1 M Tris-HCl (pH 6.8)	504 µl
	30% (w/v) Acrylamide/Bis	664 µl
	10% (w/v) SDS	40 µl
	10% (w/v) APS	40 µl
	TEMED	4 µl
Resolving gel (for 2 small gels)	Water	2.8 ml
	1.5 M Tris-HCl (pH 8.8)	1.75 ml
	30% (w/v) Acrylamide/Bis	2.3 ml
	10% (w/v) SDS	70 µl
	10% (w/v) APS	70 µl
	TEMED	2.8 µl
Protein loading buffer (2X)	Water	2.7 ml
	0.5 M Tris-HCl (pH 6.8)	1.0 ml
	Glycerol	2.0 ml
	10% (w/v) SDS	3.3 ml
	β-mercaptoethanol	0.5 ml
	0.5% (w/v) Bromophenolblue	0.5 ml
SDS-electrode buffer (10X)	Tris base	15 g
	Glycine	72 g
	SDS	5 g
	Water	ad 500 ml
Staining solution	Coomassie blue R-250	1 g
	Mathanol	500 ml
	Acetic acid	75 ml
	Water	ad 1000 ml
Destaining solution	Methanol	200 ml
	Acetic acid	76 ml
	Water	ad 1000 ml

2.2.5. Buffers for extraction and purification of His_6-tagged fusion proteins from IMAC

2.2.5.1. Buffers for BIS and BPS*

Lysis buffer	50 mM Tris-HCl 300 mM NaCl 10 mM Imidazole 10% Glycerol Water ad 500 ml
Washing buffer I	50 mM Tris-HCl 300 mM NaCl 30 mM Imidazole 10% Glycerol Water ad 500 ml
Washing buffer II	50 mM Tris-HCl 1 M NaCl 50 mM Imidazole 10% Glycerol Water ad 500 ml
Washing buffer III	50 mM Tris-HCl 300 mM NaCl 80 mM Imidazole 10% Glycerol Water ad 500 ml
Elution buffer	50 mM Tris-HCl 300 mM NaCl 250 mM Imidazole 10% Glycerol Water ad 500 ml

*pH of all buffers adjusted to 8 with HCl.

2.2.5.2. Buffers for TEV protease*

Lysis buffer	50 mM Tris-HCl 200 mM NaCl Water ad 500 ml After sonication add from autoclaved 1 M Imidazole to a final concentration of 10 Mm
Washing buffer A	50 mM Tris-HCl 300 mM NaCl 30 mM Imidazole 1 mM β-mercaptoethanol 20% (v/v) Glycerol Water ad 500 ml

Washing buffer B	50 mM Tris-HCl 300 mM NaCl 50 mM Imidazole 1 mM β-mercaptoethanol 20% (v/v) Glycerol Water ad 500 ml
Elution buffer	50 mM Tris-HCl 300 mM NaCl 300 mM Imidazole 1 mM β-mercaptoethanol 20% (v/v) Glycerol Water ad 500 ml

*pH of all buffers adjusted to 8 with HCl.

2.2.6. Buffers for desalting and SEC

For BIS	50 mM Tris-HCl 100 mM NaCl 10% Glycerol Water ad 500 ml pH 7.4 adjusted by HCl
For BPS	50 mM Tris-HCl 100 mM NaCl 10% Glycerol Water ad 500 ml pH 8.4 adjusted by HCl
For TEV protease	50 mM Tris-HCl 150 mM NaCl 1 mM DTT (Dithiothreitol) 20% (v/v) Glycerol Water ad 500 ml pH 8 adjusted by HCl

2.2.7. Buffers for IEC

Tris –HCl Glycerol	20 mM 10% pH 7.5 adjusted by HCl

2.2.8. Solution to determine the protein amount

Bradford-dye solution	Coomassie-Brilliant blue G-250	100 mg	The Coomassie-Brilliant blue G250 powder was dissolved in ethanol by stirring. Orthophosphoric acid was added, and then the volume was completed with water. The solution was filtered till no blue color could be seen. It was kept protected from light in an amber glass bottle by 4°C.
	Ethanol 96%	50 ml	
	O-phosphoric acid 85% w/v	100 ml	
	Water	ad 1000 ml	

2.2.9. Solutions for PD-10 washing and Ni-NTA agarose regeneration and recharging

PD10 washing solution NaOH (0.16 M)	Wash with 25 ml (5x column volume) then wash with water till getting a neutral eluent by testing the eluent with litmus paper
Used Ni-NTA agarose of blue color was washed with 10x volume of the following solutions in the same order	0.2 M acetic acid 30% glycerol Deionized water
Used Ni-NTA agarose of white or brown color was recharged with the following solutions in the same order	2CV 6 M guanidine HCl 5CV water 3CV 2% SDS 1CV of 25%, 50%, 75%, and 100% EtOH 1CV of 75%, 50%, 25% EtOH 1CV water 5CV of 100 mM EDTA pH 8 2CV water 2CV of 100 mM NiSO4 2CV water 2CV 6 M guanidine HCl Equilibrate with 2CV of suitable buffer Finally store the resin in 20% EtOH

3. Materials used for molecular biology

3.1. Bacteria strains

E. coli	**Purpose**	**Genotype**
DH5	Deliver high yield plasmid preparation for downstream applications, chemically competent	F'φ80δlacZ9M15endA1hsdR17(rkmk+)supE44 thi1 λgyrA96relA19 (lacZYA-argFV169) deoR
BL21(DE3) pLysS	For protein overexpression (Invitrogen), chemically competent	F-, ompT hsdSB (rB- mB-) gal dcm (DE3) pLysS (CamR)

3.2. Cloning vectors

pRSET-B vector (for protein expression)	Invitrogen
pET 52b (+) (for C-terminal SaBIS1 production)	Novagen

3.3. Enzymes

Name	**Purpose**	**Supplier**
Phusion Hot Start II High fidelity DNA polymerase	High fidelity amplification of DNA	Thermo Scientific
Dpn I	Digest the parental DNA template and to select for mutation containing synthesized DNA	Thermo Scientific
Kpn I *Nhe* I	Endonucleolytic cleavage of DNA to give specific double-stranded fragments with terminal 5'-phosphates	Thermo Scientific Thermo Scientific
*Bam*H I *Sal* I	For C-terminal His_6-tagged SaBIS1 production	Thermo Scientific Thermo Scientific
EcoR I	DNA linearization	Thermo Scientific
RNase A	Digestion of RNA for plasmid isolation	Thermo Scientific
T4 ligase	Ligation of sticky ends of DNA	Thermo Scientific
FastAP Thermosensitive Alkaline Phosphatase	Catalyzes the release of 5'- and 3'-phosphate groups from linearized vector DNA and preventing its recircularization	Thermo Scientific

3.4. Primers

All the primers were synthesized in HPSF (High Purity Salt Free) quality at Eurofins Genomics (Ebersberg, Germany).

C-terminal His-$_6$ tagged SaBIS1 primers

5 Bam Sa BIS1	Tm 60.8 °C	AGT CGG ATC CTG AGA ATC TTT ATT TTC AGG
3 Sa BIS1 TEV	Tm 80.3 °C	GCA TGT CGA CGC CCT GAA AAT AAA GAT TCT CGC ATG GAA TAG ATT CAC TAC GCA G

HaBPS and HaBPST135L mutant primers

HaBPSC167A Sen	Tm 81.7 °C	CAG GGC GCC TTC GCT GGG GGC ACG GCC
HaBPSC167A Ant	Tm 80.3 °C	C GAA GGC GCC CTG GTT GTA GAG CAT GAC GCG C
HaBPSC167S Sen	Tm 80.5 °C	CAG GGC TCC TTC GCT GGG GGC ACG GCC
HaBPSC167S Ant	Tm 79.1 °C	C GAA GGA GCC CTG GTT GTA GAG CAT GAC GCG C

MdBIS3 mutation primers

MdBIS3C167A Sen	Tm 74.6 °C	T GAA GCT GGC GCC TAT GCT GGT GCG ACA GTC C
MdBIS3C167A Ant	Tm 69.5 °C	GC ATA GGC GCC AGC TTC ATA GAT CAT GGT T C

Artificial HaBPS mutation primers

HaBPSaA133C Sen	Tm > 75 °C	AC GTG GTG TTC TGC ACG GCC TCG GGG GTC ATG
HaBPSaA133C Ant	Tm 67,1 °C	GT GCA GAA CAC CAC GTG TGT TAT TTT CGA AAT TG
HaBPSaV138F Sen	Tm > 75 °C	ACG GCC TCG GGG TTC ATG ATG CCC GGC GCA GAC TA
HaBPSaV138F Ant	Tm 67,7 °C	CAT GAA CCC CGA GGC CGT GCA G
HaBPSaY168F Sen	Tm 73,6 °C	AA GCA GGC TGT TTT GCT GGG GGG ACG GCA C
HaBPSaY168F Ant	Tm 65,4 °C	GC AAA ACA GCC TGC TTC ATA TAA CAT GAC G
HaBPSaV199M Sen	Tm 72,9 °C	A AAC ACC GCC ATG TTT TTC CAC GCC CCG AAT GAG TC
HaBPSaV199M Ant	Tm 64,8 °C	AAA CAT GGC GGT GTT TTC CGC GCA TA
HaBPSaS219A Sen	Tm > 75 °C	AA GCC CTG TTC GCA GAT GGC GCG GCT GCT CTG
HaBPSaS219A Ant	Tm 68,0 °C	TC TGC GAA CAG GGC TTG ACC CAC TAT C
HaBPSaG257A Sen	Tm > 75 °C	GT TCC GAC GGC GCC GTC ACG GCG CAT ATT TAC GAA ATG
HaBPSaG257A Ant	Tm 73,3 °C	GAC GGC GCC GTC GGA ACC CGG C
HaBPSaK272S Sen	Tm 69,5 °C	GC TAT TTT TTG AGC GAA GAT GTA ATC CCT CTG GTC C
HaBPSaK272S Ant	Tm 64,9 °C	TC GCT CAA AAA ATA GCT AAA CCC CAT TTC GTA AAT A

4. Instruments

4.1. Instruments of TUB

Instrument	Model	Company
Balance	LA 230S	Sartorius
Autoclave	Vx-120 Systec GmbH	Systec
pH Meter	Digital pH meter 325	WTW GmbH
-80°C Freezer	Hera Freeze	Heraeus
Spectrophotometer	Ultrospect 1000	Pharmacia Biotech
Magnetic rotator	VF2	IKA-Labortechnik
Ultrasonic Cell-Disrupter	Sonifier 250	Branson (G.Heinemann)
Thermo block	Dri Block BD-3	Techne
Water purification system	Arium 611 VF	Sartorius
Heating circulator water bath	TopTech MW-4	Julabo
Diaphragm vacuum pumps	VCZ 224	Ilmvac
Thermocycler	T-Professional basic gradient	Biometra
Vortex shaker	VF2	IKA
Gel documentation	Infinity-3000	Vilber Lourmat
Electrophoresis chambers:		
Agarose gels	Wide Mini-Sub Cell GT	BioRad
Polyacrylamide gels	Mini PROTEAN Tetra Cell	BioRad
Centrifuges	Universal 32R	Hettich
	Biofuge 13	Heraeus Sepatech
	AvantiA - J-E Centrifuge	Beckman Coulter
	Sigma 1-15K	Sigma Centrifuges
Incubator Shakers	Multitron	Infors HT
	KF4	Infors HT
Clean benchs	Laminar Air HLB 2472	Heraeus
	Laminar Air HLB 2460	Heraeus

Instrument	Model	Company
HPLC	LaChrom Elite System	VWR-Hitachi
Auto sampler	L-2200 auto sampler	VWR-Hitachi
Pump	L-2130 Pump	VWR-Hitachi
Detector	L-2455 diode array detector	VWR-Hitachi
Software	EZChrome Elite	Agilent
Degasser	Model 2005 Degasser	VWR-Hitachi
Column	Hyper Clone ODS column (C18, 250 x 4.6 mm, 3 µm)	Phenomenex
Column	Symmetry® glass lined column C8, 150 x 4.6 mm, 3.5 µm	Waters

4.2. Instruments of crystallization in HZI

Instrument	Model	Company
FPLC: Machine	Äkta Purifier 10	GE Healthcare
Columns	Superdex S200 16/60	GE Healthcare
	Superdex 200 10/30	GE Healthcare
ITC	VPITC	MicroCal/GE Healthcare /Malvern
MST	Monolith NT.115	Nanotemper
Thermofluor	Self-designed screen with SyPro Orange	Sigma-Aldrich ran in a BioRAD CFX96 RT-Cycler
Robot	HoneyBee 961	Digilab
Imager	Rockimager1000	Formulatrix
X-Rays: Generator	Micromax 007 HF rotating anode	Rigaku
Optics	Varimax	Rigaku
Detector	Saturn 944+ CCD	Rigaku
Goniometer	AFC-11 goniometer	Rigaku
Self-made radom screens	Design in Rockmaker Set up in Formulator34	Formulatrix

III. Methods

1. General molecular biology methods

1.1. Bacterial growth

Culturing of the *E. coli* strains was carried out as a rule in LB medium (II.2.1.) containing the appropriate antibiotics (II.2.1.). For plasmid isolation a single colony was transferred from LB plate with a sterile toothpick into 5 ml LB medium and was grown overnight at 37 ° C in a rotary shaker. To obtain the *E. coli* in sufficient quantity for protein expression, the single colony was transferred initially into 10 ml of LB medium and after 12-14 hours, 4 ml of the overnight culture were transferred into 100 ml LB medium and incubated to an OD_{600} 0.6-0.8. Since this small scale culture was not sufficient for a purpose such as crystallization larger volumes of the medium were used to express the target protein in high amount, single colony was grown in 50 ml LB medium for 6-8 hours, the flasks were placed at room temperature. In the next morning 25 ml of this culture were transferred into 400 ml of LB medium in baffle flasks and incubated 4 hours to reach OD_{600} 1. Nearly the same protocol was performed for TEV protease but differs in the growing of the bacteria first in 5 ml LB medium for overnight and in the next morning 2 ml of this culture were transferred into 400 ml LB and were growing until OD_{600} 0.6 was reached.

1.2. Enumeration of bacteria

A photometric estimation of bacterial growth and density was determined by measuring of its turbidity in the culture media (Maier, 1992) in which the light scattering is caused by the growing of the organisms and at a wavelength of 600 nm an absorbance of 1 ml of inoculated LB medium against sterile medium was measured.
Absorption of OD_{600} = 1 corresponds to an approximate *E. coli* cell count of 8×10^8.

1.3. Bacterial glycerol stock

For the permanent storage of *E. coli* BL21 (DE3) containing the desired plasmids, an aliquot of 750 µl bacterial cultures with OD_{600} 0.6 was mixed with 250 µl autoclaved stock solution (40 ml LB medium + 60 ml glycerol). The bacterial glycerol stock was stored then at -80°C. Recovery from the frozen culture was done by scratching with a sterile loop and streaking on the surface of an LB agar plate containing the appropriate antibiotic. The plate can be stored at 4° C for short-term use (max. 4 weeks).

1.4. Quantification of DNA concentration

The concentration of nucleic acids was determined by measuring the absorbance value of the DNA sample at wavelength of 260 nm. One unit absorbance value at wavelength 260 nm (ε260) corresponds 50 ng/µl double-stranded DNA (Sambrook et al., 2001).
[c] = A* ε260* dilution factor [c]; concentration of DNA.

Purity was determined by calculating the ratio of the absorbance at 260 nm to the absorbance at 280 nm. Pure samples having an A260/A280 ratio of 1.7-1.9 ensure the appropriate quality.

1.5. Agarose gel electrophoresis

Ethidium bromide and Midori green stains were used for visualization of DNA bands. They intercalate in nucleic acid molecules and its fluorescence increases 20-fold after this binding. According to the number of the samples the corresponding volume of TAE buffer and agarose weight were mixed and boiled in a microwave. After cooling down, the staining dye was added and the mixture was poured in the gel tray and left to solidify at room temperature. Samples were loaded on the gel after mixing with loading dye along with DNA ladder then the run started after putting the gel in a chamber filled with TAE buffer and under electric current (220 V). The run stopped when the visible dye in the loaded sample migrate 2/3 distance of the gel length (~30 min). Finally the bands in the gel were visualized under UV-equipped transilluminator**.**

1.6. Preparation of competent cells of *E. coli*

E.coli cells treated with calcium ions can take in the foreign DNA-plasmid and are called "competent cells". They are used in the recombinant DNA experiments for transformation; their preparation was done according to the modified methods in the literatures (Maier, 1992, Birnboim & Doly, 1979, Katsuyama et al., 2011). A single colony from an overnight agar plate or a sterile loop from a frozen culture was inoculated into 5 ml of LB medium and incubated with shaking at 37°C overnight. One ml of this fresh culture was transferred to 50 ml LB medium and incubated at 37°C with shaking until the OD_{600} reached 0.6-0.8. The culture was then chilled for 10 min on ice and the cells were harvested by centrifugation at 3000 rpm and 4°C for 10 min. The pellet was washed with 50 ml of ice-cold 0.05 M CaCl2 and the cells were harvested again by centrifugation as above and resuspended in 20 ml of 0.05 M CaCl2. The cell suspension was then kept in ice for 20 minutes. After centrifugation the pellets were resuspended in 940 µl of 0.05 M CaCl2 solution and incubated for 15 min on ice, afterwards 140 µl glycerol was added. The competent cells were stored at -80°C as 50 µl aliquots until further usage.

1.7. Transformation of plasmids into *E. coli*

1.7.1. Transformation into *E. coli* strain DH5

This strain of *E.coli* was used for plasmid maintenance because it has high transformation efficiency and high plasmid yields. The following procedure was used: 50 µl *E.coli* DH5α competent cells from -80°C storage (III.1.6.) were incubated 5 minutes in ice followed by addition of 1 µl of foreign plasmid DNA. After mixing and incubation of the mixture on ice for 25 minutes a heat shock at 42°C for 45 seconds was applied to force the foreign plasmid enter the cells followed directly by 5 minutes incubation at 4°C. In order to maximize the transformation efficiency cells were allowed to grow in 250 µl SOC medium with continuous shaking at 37°C for 60- 90 min. Finally the whole bacterial suspension is plated on LB-agar medium containing ampicillin (100 µg/ml).

1.7.2. Transformation into *E. coli* strain BL21 (DE3) pLysS

E.coli strain BL21 is mainly used for high-level of gene expression and production of recombinant protein in a bacterial system. The transformation procedure was the same as with DH5α (III.1.7.1.) except that, the heat shock time was for 20 seconds and 700-900 µl SOC medium was used for growing of bacterial culture. After one hour of incubation at 37°C about 100 µl of bacterial suspension was finally plated on LB-agar medium containing ampicillin (100 µg/ml) and chloramphenicol (30 µg/ml). Ampicillin resistance is acquired from the expression vector and chloramphenicol resistance is acquired from the host cells. For TEV protease the used host cells were BL21/DE3 tuner and only ampicillin was used for its growing. Fresh plates were used always for the desired protein before each purification and crystallization procedure.

1.8. Plasmid isolation from *E. coli* miniprep

Plasmid isolation was performed using an alkaline hydrolysis method (Birnboim & Doly, 1979) in which a single colony from the transformed DH5α plate was inoculated into 5 ml LB medium containing 100 µg/ml ampicillin and grown over night at 37°C. In the second day 2 ml cultures were centrifuged at 5000 rpm for 5 min. Bacterial pellets were then resuspended in buffer I (300 µl) containing RNase A then buffer II (300 µl) was added and the bacterial suspension was mixed by inverting continuously 6 times and incubated at room temperature for maximum 5 min. At this step lysis of the cell wall took place in addition to denaturation of large chromosomal DNA. RNA is destroyed by RNase A. Precipitation of proteins was done by adding ice-cooled buffer III (300 µl) and gentle mixing by inversion upside-down 6 times then incubation on ice for 20 min. Centrifugation at 13.000 rpm for 10 min was done to exclude the denatured proteins and genomic DNA. Then the clear supernatant containing the plasmid DNA solution (800 µl) was transferred to a new Eppendorf tube. Residual contaminants and hydrolysed protein were extracted by vortex with 800 µl chloroform followed by centrifugation at 13.000 rpm for 10 min. The aqueous layer was then transferred to a new Eppendorf tube and isopropanol (0.7 volume of the aqueous layer volume) was added then the mixture was centrifuged at 13000 rpm for 10 min to precipitate plasmid DNA. The pellets were washed with 70% ethanol (500 µl) followed by centrifugation at 13.000 rpm for 10 min. The supernatant is discarded and the plasmid pellets are dried by 37°C and then dissolved in 50 µl distilled water. This purified plasmid can be directly used after measuring the concentration as described in (III.1.4.) for transformation (III.1.7.) or stored at -20°C for further usage.

1.9. Digestion with restriction endonucleases

It is possible to cut a specific region of DNA strand using endonucleases. These enzymes hydrolyze the phosphodiester bond between the 5'-phosphate group of a nucleotide and the 3'-hydroxy group of the neighboring nucleotide by the recognition of a palindromic sequence which is unique for each restriction endonuclease. This will help in the ligation of DNA insert and vector which are cut with the same endonuclease. In 20 µl reaction mixture about 0.5-1µg plasmid DNA was incubated at 37°C for 3 hours with 1 µl endonuclease and 2 µl of 10x restriction endonuclease specific buffer.

1.10. Isolation of DNA by gel extraction

By the use of innuPREP DOUBLE pure kit from Analytikjena. Salts and macromolecules are removed by washing with ethanolic wash buffer without elution of DNA fragments. With an alkaline buffer of low ionic strength, the DNA is eluted and stored at -20°C.

1.11. Ligation and dephosphorylation

The DNA insert and the vector were mixed in molar ratio of 3:1, added to 1 µl of 10x T4 DNA ligase buffer, mixed with 0.25 µl of T4 DNA ligase, the final volume was completed with distilled water to 10 µl. The mixture was incubated for 3hrs at 16°C and subsequently transformed in *E.coli* DH5α as described in (III.1.7.1.). Dephosphorylation of the vector after digestion is an important step before the ligation because it minimizes the ligation of the digested vector with itself. This process will help in the transformation of more recombinant but not recircularized vector plasmid to the bacterial cells. In this reaction 2 µl of 10x reaction buffer for AP (alkaline phosphatase) was mixed with 1 µl FastAP Thermosensitive Alkaline Phosphatase and the final volume was completed to 20 µl using the digested vector. The mixture was then incubated for 20 min at 37°C followed by inactivation of the enzyme for 15 min at 65°C. Finally the vector was purified from the reaction components as described in (III.1.10.).

1.12. Polymerase chain reaction (PCR)

This reaction utilizes DNA polymerase which synthesizes complementary strands of DNA using a DNA template and primers by adding nucleotides to the 3′-end of the primer. It requires three steps:
The denaturation step: the separation of the double strands to form single stranded DNA by heating the DNA for 20–30 s at 95-98 °C.
The annealing step: the primers anneal to their specific template by using specific annealing temperature (between 45-70 °C) which is 3-5 °C below the Tm of the primers.
The extension step: the multiplication of the initital DNA template by heating at 72 °C which is the optimum temperature for most thermostable DNA polymerases.

The reaction components and conditions were as follows using the recommended protocol of Q5® High-Fidelity DNA Polymerase:

Component	Amounts
5X Q5 Reaction Buffer	10 µl
10 mM dNTPs	1 µl
10 µM Forward Primer	2.5 µl
10 µM Reverse Primer	2.5 µl
Template DNA	100 ng
Q5 High-Fidelity DNA Polymerase	0.5 µl
Nuclease-Free Water	to 50 µl

1.13. Site-directed mutagenesis PCR (SDM-PCR)

The QuickChange™ site-directed mutagenesis procedure from Stratagene was used, in which two oligonucleotide primers with the desired mutation were designed according to the modified primer protocol of Liu and Naismith (2008). The primers will anneal to the template DNA using PCR and after extension by DNA polymerase a mutated plasmid will be generated which can be used as a template in the subsequent cycles.The conditions were as follows:

PCR reaction constituents		**PCR program**		
10 pmol sense mutagenic primer	2.5 µl	**Temp °C**	**Time**	**Step**
10 pmol Anti-sense mutagenic primer	2.5 µl	Lid 98		Preheating on
Template DNA 5-50 ng/µl	1 µl	98	Pause	
5x Phusion HF buffer	10 µl	98	30 sec	Initial denaturation
2 U/µl Phusion Hot Start II DNA Polymerase	0.5 µl	98	10 sec	Denaturation
		63	30 sec	Anneling
10 mM dNTPs mix	1 µl	72	2 min	Extension
Up to 50 µl PCR water		72	10 min	Final extension
		4	Pause	

25 cycles

One µl of *Dpn*I restriction enzyme was added to the SDM-PCR product and the mixture was incubated for 3hrs at 37°C. This enzyme is specific for methylated and hemimethylated DNA and it will digest only the parental DNA template. After that 6µl of DpnI treated PCR product was used for transformation of *E. coli* DH5α competent cells (III.1.7.1.).

1.14. Expression and protein induction

Protein expression was performed by low temperature to slow the metabolic processing of the bacteria and then giving the opportunity for the protein to be correctly folded. After growing of the bacteria in 400 ml LB medium until OD_{600} 1, the expression was then induced by the addition of 0.1-0.2 mM IPTG to the bacterial suspension and the flasks were incubated at 24°C in the shaker for 4 hours. Thereafter the cells were harvested by centrifugation at 6000 rpm for 16 min. At this point the cell pellets were stored at -20°C for several weeks. For TEV protease, the induction was performed by the addition of 0.2 mM IPTG and the incubation was at 20°C for overnight.

1.15. Cell lysis

The bacterial pellet from 800ml medium (III.1.13.) was suspended in 16-18 ml lysis buffer and disrupted by sonication on ice for 8-10 min in the sonifier at duty cycle 40% and output control of 1.5. This solution was centrifuged at 4°C for 20 minutes at 14,000 rpm and the clear supernatant was used for protein purification by affinity chromatography. For TEV protease, a pellet from 2 L medium was suspended using 40 ml lysis buffer and after sonication the solution was centrifuged at 4°C for 20 minutes at 20,000 rpm.

2. General biochemical methods

2.1. Protein purification

2.1.1. IMAC (immobilized metal affinity chromatography) (Ni-NTA agarose)

Since all recombinant proteins used in this work were cloned with a polyhistidinetag (His_6- tag), they could quickly and easily purified by the use Ni-NTA agarose (Quiagen, Hilden). The Ni2 + ions have six complex binding sites, four sites are bound to nitrilotriacetic acid which is linked to sepharose matrix and the other two remaining binding sites are available to interact with each histidine of the His_6 -tag. The elution of the protein can be achieved either through a reduction of pH (pH 4.5-5.3) as the protonated nitrogen can no longer bind to the nickel ions or by the addition of imidazole to the elution buffer because imidazole competes with histidin residues due to its structural equivalence. At concentrations between 100-250 mM it can displace the His_6-tag which is bound to the matrix. Since the purification under native condition was carried out an increase in the concentration of imidazole was gentler on the protein as pH reduction. The addition of low concentrations of imidazole for washing and lysis buffer was to prevent the nonspecific binding of foreign proteins. The solution (III.1.14.) obtained after cell disruption was added to 0.8 ml Ni-NTA agarose and gently mixed for one hour at 4°C. After this incubation the solution was poured through an empty PD10 column and the gel matrix was retained. The agarose with the bound protein was washed using the wash buffers in (2.2.5.).

Firstly it was washed 10 times with 1 ml wash buffer I (30 mM imidazole), then followed by 6 times with 1 ml wash buffer II (50mM immidazole) and finally by 4 times with 1 ml wash buffer III (80 mM immidazole). The elution of the target protein was performed with 5 times 0.5 ml elution buffer (250 mM imidazole). For TEV protease, 2 ml Ni-NTA agarose was used and the purification was performed by washing the column 10 times with 1 ml wash buffer A then followed by 8 times with 1 ml wash buffer B.

For large amounts of protein, the required amount of Ni-NTA agarose was calculated according to the column capacity in which 1 ml Ni-NTA agarose can bind 5-10 mg protein. The washing times should be increased to 10 times for different washing buffers.

2.1.2. Size exclusion chromatography (SEC)

2.1.2.1. PD-10 columns

Although the imidazole in the buffers did not affect the enzyme activity, it should be removed for crystallization approaches and storage of the protein at -20°C. It was also necessary to reduce the ionic strength during the Ni-NTA purifications for the crystallization. Therefore the protein solution obtained in (III.2.1.1.) was desalted using size exclusion chromatography. For this purpose, PD10 column (3.5 ml bed volume, separation region 1000-5000 Da Amersham Biosciences) was equilibrated with 30 ml of the desalting buffer (II.2.2.6.) and then 2.5 ml of the protein solution from (III.2.1.1.) was applied. When the solution is completely immersed into the separation bed, elution of the protein was carried out with 3.5 ml of the desalting buffer, so the low molecular weight salts will be retained within the pores of the gel matrix and the protein can be eluted with the void volume.

2.1.2.2. Superdex high resolution columns

As PD-10 columns were used for small protein amounts highly concentrated samples of protein were desalted and highly purified using preparative size exclusion chromatography columns such as superdex S200 10/30 for small amount samples (less than 5mg/lml) and superdex S200 16/60 for large amount samples (more than 10 mg/ml). Moreover the use of such columns was essential to identify the homogeneity of a protein sample before the crystallization process. As SEC separations are based on diffusion into and out of the pores of the column's packing material, the larger proteins cannot access the pores and elute earlier such as aggregates that may formed during any stage of the purification process. The smaller proteins are trapped within the pores, which results in longer retention times and later elution, such as our target proteins. After equilibration of the column using the desalting buffer, 0.5-1 ml of highly concentrated protein samples (10-30 mg/ml) were injected and the eluted fractions corresponding to the sharp large peak were collected and concentrated to the desired concentration for crystallization screens.

2.1.3. Ion exchange chromatography (IEC)

To get protein in a good quality especially for crystallization purposes, different purification steps are needed. Affinity chromatography and size exclusion chromatography are the most famous techniquese and they work for almost all proteins such as BPS. But for other proteins such as BIS additional steps should be included to get a sample in a good quality, which is ion exchange chromatography.

2.1.3.1. Mono Q anion exchange column

Anion exchange chromatography was used to purify substances with a negative charge at high pH level, such as proteins. The column is pre-equilibrated with the buffer (II.2.2.7.) before the protein mixture is applied, then the elution can start using a salt gradient (0-1 M) NaCl. The negative ions in the salt solution (Cl-) compete with proteins in binding to the resin. Unbound proteins are collected in the flow-through and/or in subsequent buffer washes, but proteins that bind to the resin are retained and can be eluted by gradual increase in the salt concentration in the elution buffer.

2.2. Cleavage of His_6-tag using TEV protease

TEV protease (Tabaco-Etch-Virus) is an enzyme that was used to cut the His_6-tags from the His_6-tagged proteins. Since the enzyme itself has a His_6-tag, it binds easily to be removed from the final purified protein solution. For the use of TEV protease the proteins had to be provided with a recognition site for this enzyme directly in the beginning of the protein that was cloned. This recognition site has seven amino acids Glu-Asn-Leu-Tyr-Phe-Gln-(Gly/Ser) (ENLYFQ (G/S), where the cleavage occurs between the Gln and Gly/Ser residues. The available construct with this site was provided for *Sorbus aucoparia* BIS1 by Dr. Maren Lütge (Lütge, 2012). The TEV protease construct that was used in this work was provided by HZI (Helmholtz Centre for Infection Research**)** in Braunschweig and was purified and stored at -80°C in our lab. The concentrated proteins were cut with TEV at different temperatures, 4°C, 16°C, 25°C and 37°C, with gentle shaking for different time intervals. Different ratios of TEV and the target proteins were used ranging from 1:50, 1:30 and 1:10. After cutting, the TEV protease and His_6 were removed via affinity chromatography with Ni-NTA agarose (III.2.1.1.). The target proteins were then collected in the flow through and the washing fractions.

2.3. Concentration of protein for crystallization

For the crystallization approach, the protein concentration should be high (about 20 mg/ml) at the end of the purification processes. For this purpose, the protein solution from III.2.1 was concentrated using vivaspin centrifugal concentrators (Sartorius, Göttingen). The membrane filter of these concentrators had a molecular weight cut-off of 10 kDa whereby proteins in size of 86 kDa are retained in the solution above the membrane. The centrifugation was performed at 9000 rpm and 4°C for 10-60 min to reach the final concentration needed for gel filtration or crystallization.

2.4. Determination of protein content

Colorimetric measurement of protein was carried out using the Bradford protein assay (Bradford, 1976) in which a coomassie brilliant blue G-250 dye interacts with aromatic and basic amino acids of the protein. The negative charged sulfonate anions of the dye donate charges to the cationic amino acids, which results in the exposure of the hydrophobic part of the protein that will interact with the phenyl residues of the dye by van der Waals forces. Upon interaction with protein the absorption maximum of the dye shifts from 465 nm (red, free form) to 595 nm (blue, bound form). The assay was done using 900 µl Bradford solution that was completed to 1000 µl by the addition of 100 µl of protein and buffer mixture. After gradual mixing a blue color appears and reaches to its maximum after two minutes and is stable for 1 h. For concentration measurement, a standard curve was constructed using the absorbance of serially diluted samples of bovine serum albumin (BSA) solution (1 mg/ml).

2.5. Protein storage

For enzyme assays, it was possible to store aliquots of the protein solution at 4°C for about a month or more without any effect on the activity, especially in the crystallization buffers which contain glycerol and no imidazole. However, at -20°C the activity decreased gradually depending on the freezing and thawing times. Enzyme batches that were used for the crystallization approaches were fresh in most cases, but it was also possible to store the protein at -80C° in liquid nitrogen for a short period until the time of the next crystallization screen.

2.6. SDS gel electrophoresis (SDS-PAGE)

SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) is the most common analytical technique used to separate and characterize proteins according to their molecular masses(Laemmli, 1970) and it was mainly used in this work to detect the size and the purity of the proteins. In the stacking zone proteins migrate through large pores; no separation takes place rather all the proteins in a sample will collect as a thin compact band at front of the resolving gel. The resolving gel has a narrower pore size and proteins will be separated according to their masses, the low molecular weight proteins will run faster than large molecular weight proteins. Bisacrylamide interconnect the polymer chains of acrylamide and the polymerization reaction is initiated by the addition of ammonium peroxydisulfate (APS), a radical initiator, and N,N,N′,N′-tetramethylethylenediamine (TEMED), a polymerization catalyst. The degree of polymerization and hence the pore size of the gel is dependent on the amount of acrylamide and the pH of the polymerization reaction.

2.6.1. Sample preparation

The protein samples were denaturated in the presence of the loading buffer (II.2.2.4.) at 95°C for 5 min. During denaturation the protein is defolded and becomes uniformly coated with negative charges from SDS in the loading buffer, this process will cancel the differences between proteins regarding their charges. Glycerol in the sample buffer ensures that all samples will sink and settle down in the pockets of the SDS-PAGE gel, while bromophenol blue, a small size dye, migrates faster than any protein and helps trace the protein bands in the gel.

2.6.2. Setting of the gel

The resolving gel ingredients (II.2.2.4.) are mixed together then poured between the glass plates of the instrument and overlaid with water to avoid interaction with air. After gel formation (20 min), water is absorbed away with filter paper. The same procedures are carried out for the stacking gel, except that a comb is added to create the sample pockets. After 20 min or less the gel is formed.

2.6.3. Electrophoresis and protein band detection

The gel is immersed in a chamber filled with the electrophoresie buffer (II.2.2.4), so the sample pockets are washed to remove any air bubbles that may hinder the electrophoresis. Protein samples are prepared as mentioned before (2.5.1.) and loaded in the pockets of the stacking gel. The gel was run at 35 mA / 200 V max supplied by a standard power pack P25 (Biometra). The run ends when the dye bromophenol blue reaches the front of the gel. After electrophoresis the protein bands are detected by immersing the gel in coomassie blue staining solution (II.2.2.4.) with gentle shaking for 10-30 min and then de-stained by immersing the gel in methanol-acetic acid de-staining solution (II.2.2.4.) for 4 hrs or overnight.

3. General analytical methods

3.1. HPLC and enzyme assay

A standard assay was followed up for all enzymes in this work to ensure the activity of the protein after purification and before the crystallization process. This protocol was optimized to detect the activity for most type III PKSs in our research group (Liu et al., 2003). The reaction mixture was prepared using the listed amounts of enzyme and substrates in the following table. The final volume was completed to 250 µl using the assay buffer (II.2.2.2.). Then incubation at 35°C for 10 min was carried out.

The reaction was stopped by further addition of 27 µl 50% acetic acid and the enzymatic products were extracted twice with 250 µl ethyl acetate by vigorous mixing and then centrifugation at 13.000 rpm for 10 min. The two organic phases were combined and dried under vacuum for 30 min, the final residue was dissolved in 50 µl of methanol for HPLC analysis.

	Volume	End concentration
Enzyme		2-4 µg
Benzoyl-CoA (230.5 µM)	15 µl	13.8 µM
Malonyl-CoA (234.5 µM)	20 µl	18.8 µM
0.1M KH2PO4 buffer (pH 7) ad to 250 µl		

4. Protein crystallization

Protein concentrations can be enhanced to form crystals when the solution in which they are dissolved becomes supersaturated. As many methods in protein crystallization were developed to promote this super-saturation stage the most abundant and the one that was used in this work is the vapour diffusion method.

4.1. Vapour diffusion

A drop of the highly concentrated soluble protein solution, which was mixed equivalently with the reservoir contents is allowed to equilibrate against the reservoir containing the crystallization components (buffers and precipitants). After equilibration the concentration of the protein and the precipitants in the drop will increase and thus prompt the growth of the crystal (Russo Krauss et al., 2013).

4.1.1. Sitting drop method

In this method a drop consisting of equal amounts of protein and buffer solution was placed on a platform over a well, which contains a higher volume of the same buffer solution (Figure III.1). Random screens were done in the first crystallization process of each protein using a robot, which pipettes 60 µl of many buffer solutions from different kits into a plate containing 96 wells. Thereafter, 0.2 µl was taken from the well to be placed on the platform and then 0.2µl of the highly concentrated protein (5-28 mg) was added. Finally, the plates were sealed with a permeable film and kept at R.T under regular checking by taking images periodically using a rockimager. The kits involved different buffers, precipitants and pH values to get the first hits, which developed even small disordered crystals. Then, further optimization of the promising conditions was done using a formulator device and new screens were designed to enhance the growth of good crystals.

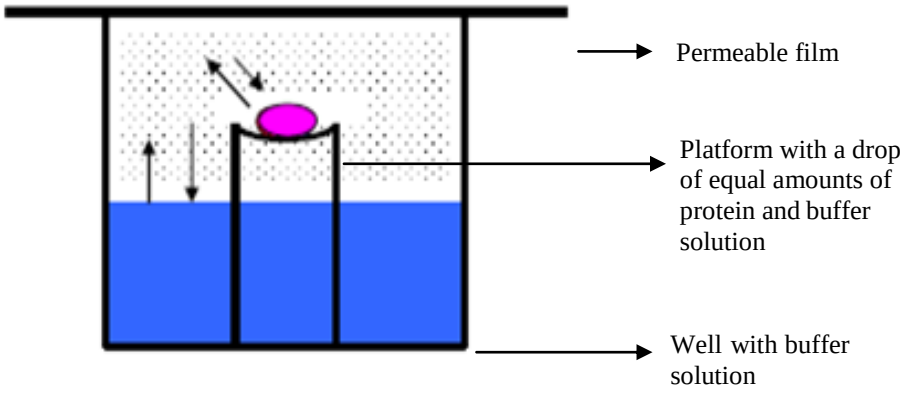


Fig III.1 Sitting drop method.

4.1.2. Hanging drop method

This technique is not quite different from the previous method, except that the drop will suspend from a cover slip over the well (Figure III.2). It was used to obtain larger protein crystals and to optimize the crystallization conditions. By the variation of salt concentrations, pH values, buffer concentrations and the addition of different precipitants, an improvement in protein crystallization should be achieved. In this method, plates of 24 crystallization wells were used, into which 700 µl each of the respective crystallization buffer were pipetted. Then 2.0 - 5.0 µl of this buffer solution were mixed with the same volume of the protein solution and applied to the inner part of the screw cap. The plates were finally closed with the caps and kept at R.T under regular checking.

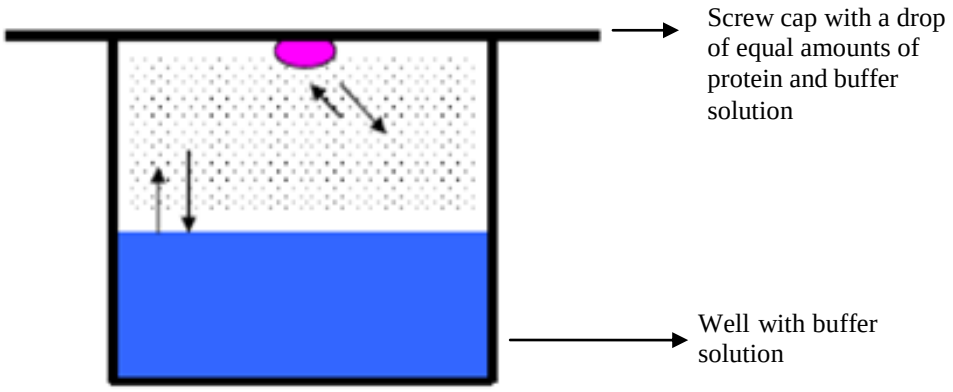


Fig III.2 Hanging drop method.

4.2. Seeding

Seeding is a tool used to improve crystal growth and formation. In this technique, previously grown crystals are washed with water. Either a single crystal was transferred and applied into a new protein drop containing fresh buffer solution (macroseeding) or too small crystals that were obtained after smashing an old crystal were transferred and applied into a new protein drop containing fresh buffer solution (microseeding) (Figure III.3). Seeding can also be used as a technique for optimization of the crystallization conditions (Bergfors, 2003).

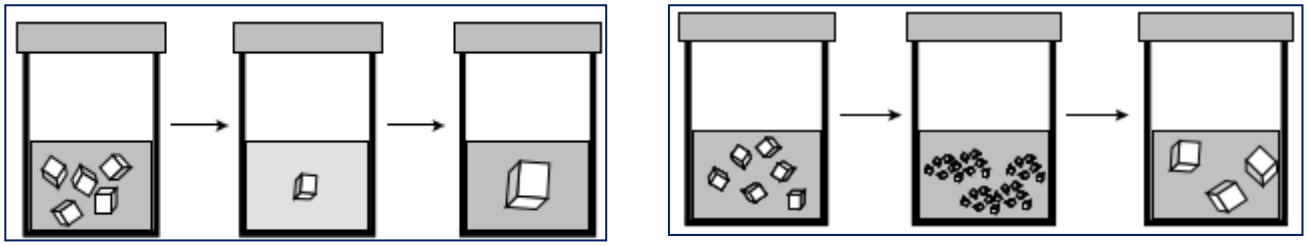


Fig III.3 Macroseeding (left) and microseeding (right).

4.3. Soaking and co-crystallization

Getting the structure of enzyme-substrate or enzyme-product complexes is one of the important aims of this project. It provides us with more information about the binding pattern of these molecules in the enzymes and determines the essential residues in the active site cavity, which are important for enzyme interaction and activity.

Soaking is the method of choice because of its ease. In this method, the growing crystal of the enzyme was soaked for a period of time with the substrates or the products. Different time intervals were used ranging from 10 sec, 5 min, 2-3 days to 2 weeks. Also, different concentrations of the substrates and the products were used starting from 10 mM until 20 mM.

Co-crystallization is different from soaking. The protein solution is mixed with the substrate or the product solution and allowed to incubate on ice for a period of time. Thereafter, the mix was subjected to the crystallizations screens. For co-crystallization we also used substrates or products from 10 mM until 20 mM and the incubation time was 1 hr.

4.4. Thermofluor assay

This is a thermal shift assay, which measures the unfolding pattern of a protein. It measures the binding affinity of the protein to different ligands and additives. This binding is evaluated from the difference of the unfolding temperature (ΔT_m) in the presence of ligands relative to the one in the absence of ligands. A sensitive fluorescent dye is used to detect the thermal unfolding of the protein after binding (Lo et al., 2004). When the thermal shift increases, then the protein is more unfolded in response to the ligand which was added to the protein solution. However, when the thermal shift decreases then the protein is less unfolded in response to the ligand which was added to the protein solution. This observation led to the use of the Thermofluor assay in the improvement of protein stability by the addition of some substances like salts to the solution.

Thermofluor Manual

A stock of 5000x Sypro Orange was diluted 100 times to get 250 µl of 50x dye solution. About 250 µl protein solution of a 2 mg/ml concentration was also prepared.

Buffer screen

Masterblock of different buffers (A, B, C and D) was used, in which the following amounts were pipeted in a 96-well PCR plate.

40 µl of the buffer
5 µl of the protein
5 µl of Sypro Orange

The components were mixed by pipeting up and down. Then the seal film was used to close the plate. After centrifugation of the plate to get rid of air bubbles, it was put into the PCR machine and the lid was closed. The run was started using a specific CFX program. Finally, the data were stored.

Additive screen

After the buffer screen, one buffer condition was chosen to test different additives (F, G and H). We prepared the buffer of interest as 8x concentrated and pipetted as follows, using the same previous concentrations of the protein and Sypro Orange.

35 µl of the additive solution
5 µl of the 8x buffer
5 µl of the protein
5 µl of Sypro Orange

After mixing by pipetting the screen was performed as previously described.

	1	2	3	4	5	6	7	8	9	10	11	12
A	50mM Acetate pH 4.0	50mM Acetate pH 4.4	50mM Citrate pH 5.0	50mM Citrate pH 5.4	50mM MES pH 6.0	50mM MES pH 6.4	50mM HEPES pH 7.0	50mM HEPES pH 7.4	50mM Tris-HCl pH 8.0	50mM Tris-HCl pH 8.4	50mM CHES pH 9.0	50mM CHES pH 9.4
B	50mM Acetate pH 4.0 100mM NaCl	50mM Acetate pH 4.4 100mM NaCl	50mM Citrate pH 5.0 100mM NaCl	50mM Citrate pH 5.4 100mM NaCl	50mM MES pH 6.0 100mM NaCl	50mM MES pH 6.4 100mM NaCl	50mM HEPES pH 7.0 100mM NaCl	50mM HEPES pH 7.4 100mM NaCl	50mM Tris-HCl pH 8.0 100mM NaCl	50mM Tris-HCl pH 8.4 100mM NaCl	50mM CHES pH 9.0 100mM NaCl	50mM CHES pH 9.4 100mM NaCl
C	50mM Acetate pH 4.0 200mM NaCl	50mM Acetate pH 4.4 200mM NaCl	50mM Citrate pH 5.0 200mM NaCl	50mM Citrate pH 5.4 200mM NaCl	50mM MES pH 6.0 200mM NaCl	50mM MES pH 6.4 200mM NaCl	50mM HEPES pH 7.0 200mM NaCl	50mM HEPES pH 7.4 200mM NaCl	50mM Tris-HCl pH 8.0 200mM NaCl	50mM Tris-HCl pH 8.4 200mM NaCl	50mM CHES pH 9.0 200mM NaCl	50mM CHES pH 9.4 200mM NaCl
D	50mM Acetate pH 4.0 500mM NaCl	50mM Acetate pH 4.4 500mM NaCl	50mM Citrate pH 5.0 500mM NaCl	50mM Citrate pH 5.4 500mM NaCl	50mM MES pH 6.0 500mM NaCl	50mM MES pH 6.4 500mM NaCl	50mM HEPES pH 7.0 500mM NaCl	50mM HEPES pH 7.4 500mM NaCl	50mM Tris-HCl pH 8.0 500mM NaCl	50mM Tris-HCl pH 8.4 500mM NaCl	50mM CHES pH 9.0 500mM NaCl	50mM CHES pH 9.4 500mM NaCl
E												
F	2mM ATP	2mM ADP	2mM AMP	2mM GTP	2mM GDP	2mM NAD	2mM FAD	1% Glycerol	5% Glycerol	10% Glycerol	20% Glycerol	3% DMSO
G	1mM DTT	5mM DTT	10mM $CaCl_2$	10mM $MgCl_2$	10mM $MnCl_2$	10mM $ZnCl_2$	10mM $NiCl_2$	10mM $FeCl_3$	100mM KCl	100mM LiCl	200mM KSCN	10mM spermid-in
H	10mM Arginin	10mM Alanin	10mM Methion-in	100mM Glycin	3% Glucose	3% Galactose	10mM Urea	5% PEG400	0.5% n-hexyl gluco-pyranosid	0.05% n-decyl gluco-pyranosid	0.01% Triton X100	H_2O as referen-ce

4.5. Isothermal Titration Calorimetry (ITC)

This is another technique which estimates the binding interaction between macromolecules and ligands by measuring all thermodynamic parameters needed for binding such as affinity, enthalpy and stoichiometry. The experiment involves the titration of a specific ligand into the protein under isothermal conditions. The produced thermal signal after titration measures the heat absorbed or released after binding between the ligand and the protein. After many injections of ligand, this signal will reach zero as the protein is fully saturated by the ligand (Figure III.4) (Duff et al., 2011).

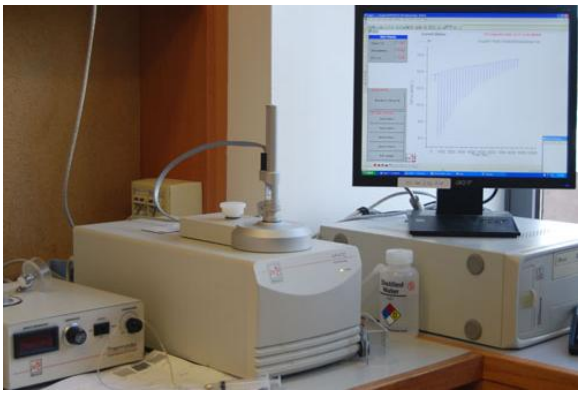


Fig III.4 Isothermal calorimetry setup.

ITC Manual

The preparation of 2 L degassed distilled water and 1L degassed buffer is essential before starting the experiment to avoid any noise signals from the bubbles in the solutions. After dilution of the ligand and the protein with the buffer, they were also degassed using Thermovac.

1. After washing the reference cell, it is filled with 2 ml degassed water.
2. After washing the sample cell several times with water, it is filled with 2 ml degassed buffer solution in addition to 2 ml degassed protein solution.
3. The syringe is also washed with the degassed water and then with the degassed buffer. The needle is dried with tissue to remove any adhesive drops.
4. The syringe is filled with the degassed ligand and placed in the sample cell.

After setting the previous conditions, the ITC experiment starts and the essential parameters are fixed using a specific computer program.

1. Mostly fewer injections with higher injection volumes are used (8-10 µl) - (30 injections).
2. The time spacing between the injections is 240 sec. The initial delay before the first injection (120 sec) is important for baseline establishment.
3. The temperature is mostly adjusted to room temperature but a range of 20-37C° can be used.

After finishing the experiment, the data are collected and the results are analysed.

4.6. X ray measurement

The protein crystals were taken from the stock solution using a Nylon-eyelet and washed in cryo-buffer. They were taken again with the loop and stored in a dry cold nitrogen stream. The addition of glycerol prevents the formation of ice crystals, which may interfere later with the image. To avoid thawing during the measurement, the crystals were washed again with cold nitrogen. The generator MicroMax 007 was used as a source for X-ray radiation and the diffraction pattern of the crystals was recorded and stored using a Saturn 944+ CCD detector. The exposure time for measurement was 2 min and complete datasets were recorded after rotation of the crystals in 0.5° angular steps in the range of 0-180° (Figure III.5).

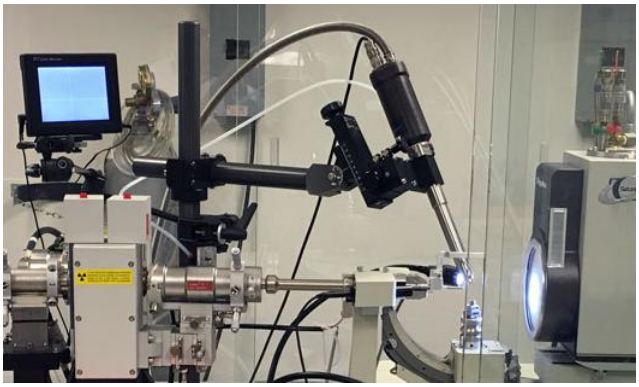


Fig III.5 X-ray radiation system.

4.7. Data analysis and structure elucidation

Many datasets and images were obtained after x-ray diffraction. At this stage, the crystallographer should follow many steps for data processing to get the final structure of the enzyme, as follows.

1. **Indexing** of the reflections (spots in the diffraction pattern) to determine the adequate space group (symmetry of the crystal).
2. **Merging and scaling** of the reflections by detection of the relative intensities and strengths of the spots in different images.
3. **Integration** of the data considering the determined space group by getting only one file with the records of indexes and intensities of all reflections in the images.
4. **Importing** of the integrated data using a data processing program (PHENIX, CCP4…etc)
5. **Phasing** of the data and building of an electron density map by molecular replacement with known related structures to determine the orientation and the position of the molecules in the crystal and to find the molecular model.
6. **Refining** the molecular model and depositing it to pdb.

Schematic diagram for data analysis and structure elucidation

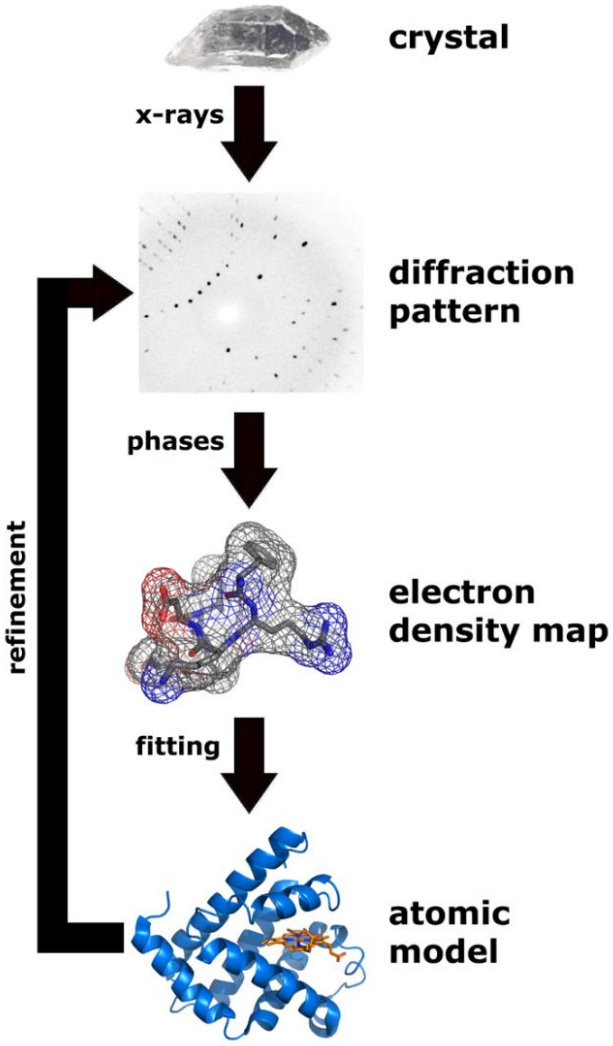


Fig III.6 Data analysis and the steps of structure elucidation.

IV. Results

In 2007, Torben Raeth was able to crystallize BIS1 from *Sorbus aucuparia* (SaBIS1) without His_6-tag (Raeth, 2007). Two datasets at 20.00–3.15 Å and 22.88–3.16 Å were used for structure refinement, based on the crystal structure of alfalfa CHS2. The crystals were small, which explains their low resolution, and the data were not enough to elucidate the aldol cyclization mechanism of SaBIS1. The structure also showed unusual electron density close to the catalytic Cys because of the presence of oxygen due to the oxidation of the sulphur (Raeth, 2007). In 2012, Maren Lütge continued the crystallization of SaBIS1 even with different constructs, but the crystals were small in size and not enough for x-ray measurement (Lütge, 2012). In this work, we basically started from the beginning by optimization of expression, purification, His_6-tag cleavage, protein stability and homogeneity tests in addition to buffer screening for crystallization.

1. TEV protease purification

The TEV construct provided by HZI, Braunschweig, contained a His_6-tag, allowing for purification on Ni-NTA-agarose. One to 2 mg of TEV protease (~ 35 kDa) were produced using 5 L LB medium and an induction time of 20 h. The pellets were harvested and subjected to affinity chromatography (III.2.1.1.). The eluate was collected and buffer exchange was carried out using PD-10 (III.2.1.2.1.). The protein content was determined using the Bradford method (III.2.3.). The fractions were supplied with 50% (v/v) glycerol and stored at -80°C. Aliquots were used to examine the purity by SDS-PAGE (III.2.5) (Figure IV.1). The TEV purification process had to be fast as this protease has the property of autolysis, i.e. cleaving itself, resulting in reduction of its efficiency in target protein cleavage.

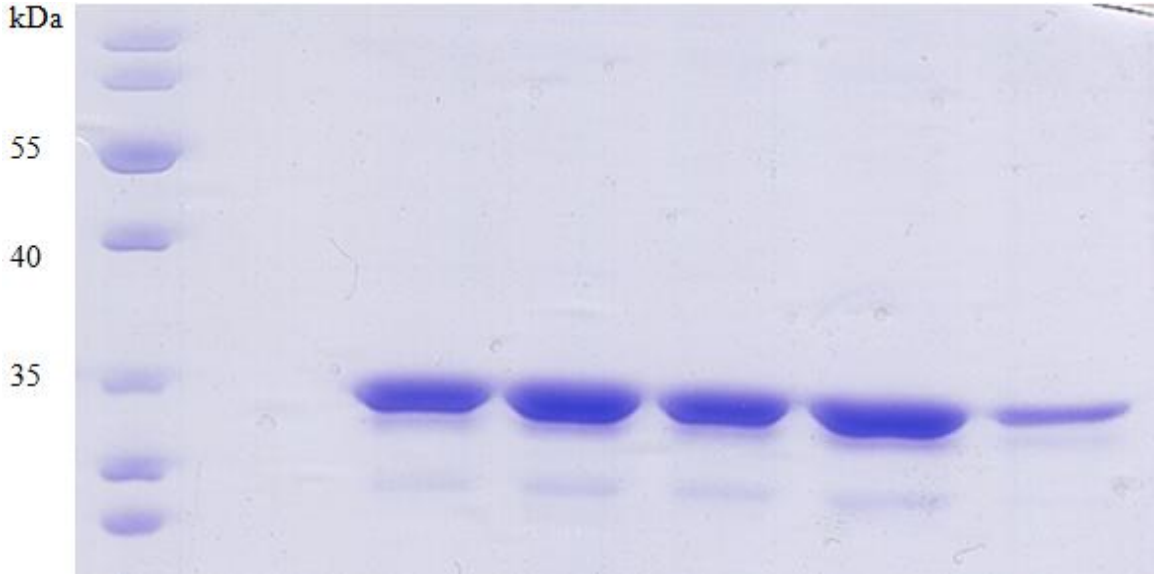


Fig IV.1 SDS-PAGE for checking the purity of TEV protease. Five fractions were tested.

2. Cutting and removal of His_6-tag

2.1. N-terminally His_6-tagged SaBIS1

The optimization of the cleavage with TEV prortease was performed using N-terminally His_6-tagged SaBIS1 as the target protein. The best known working conditions for TEV protease were cleavage at 30°C using a 1:50 ratio of TEV to target protein. In the beginning, cutting was carried out using the 1:50 ratio at 4°C (Lütge, 2012). However, at 4°C the cleavage was slow and incomplete and a double band appeared at almost all time intervals (Figure IV.2).

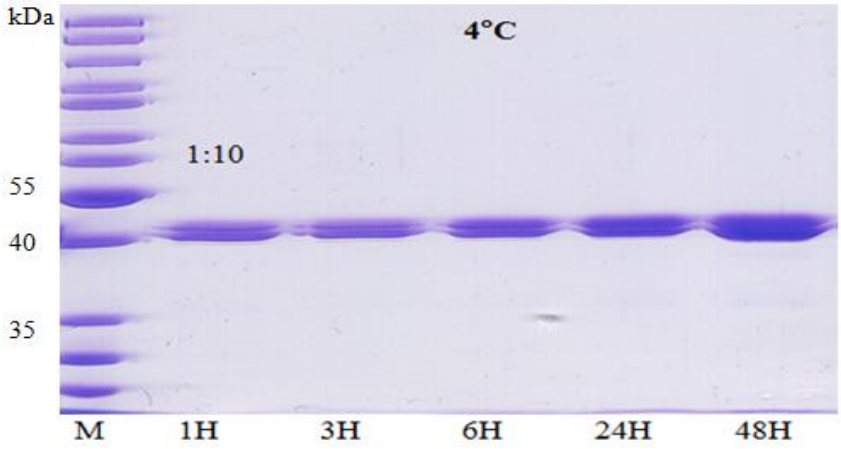


Fig IV.2 TEV protease cleavage of SaBIS1 at 4 °C. SDS PAGE shows incomplete cutting of tagged SaBIS1 using a 1:10 ratio of TEV protease and target protein at 4 °C.

Different ratios and temperatures were tried to achieve a complete cutting in short time intervals without influence on the activity of the target protein beside the use of less amounts of TEV protease. At 16 °C, when the ratio of TEV protease to SaBIS1 was 1:30, it needed 20 h to reach the complete cutting. When the ratio of TEV protease to SaBIS1 was 1:10, only 6 h were needed for the complete cleavage (Figure IV.3a).

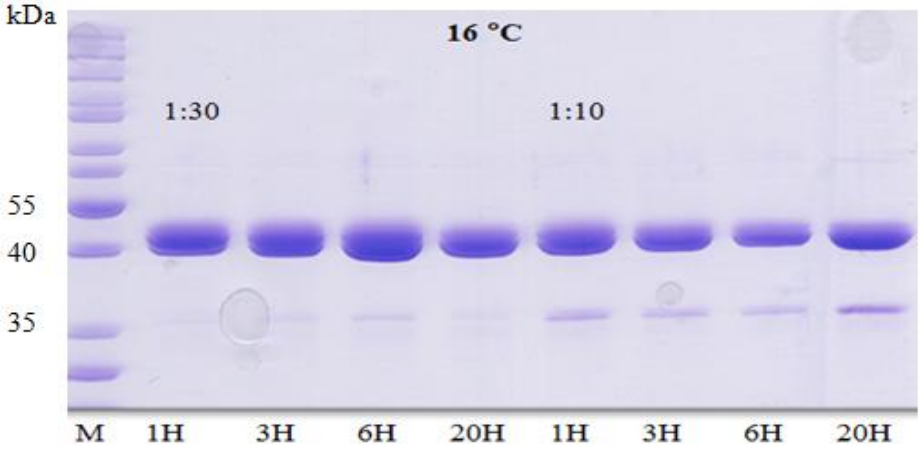


Fig IV.3a Cleavage of tagged SaBIS1 at 16 °C using TEV protease and varying hours. The SDS-PAGE gel shows efficient cutting of tagged SaBIS1 at different ratios of TEV protease to SaBIS1. The faint lower band (~ 35 kDa) represents TEV protease.

Similar results were obtained at 25 °C. Efficient cutting was observed after 20 h using a 1:30 ratio and nearly complete cutting was achieved after 6 h using a 1:10 ratio (Figure IV.3b).

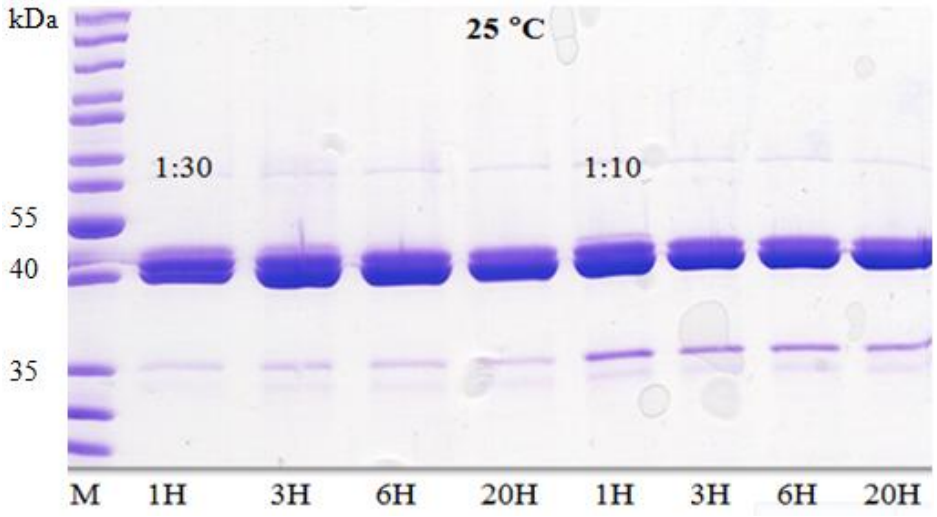


Fig IV.3b Cleavage of tagged SaBIS1 at 25 °C using TEV protease at various ratios and times. The SDS-PAGE gel shows the efficient cutting of tagged SaBIS1 after 6 h using a 1:10 ratio of TEV protease to SaBIS1. The faint lower band (approx. 35 kDa) represents TEV protease.

A faster and efficient cleavage was carried out by using a 1:30 ratio of TEV protease to the target protein and incubating the mixture under slow shaking at 37 °C (Figure IV.3c).

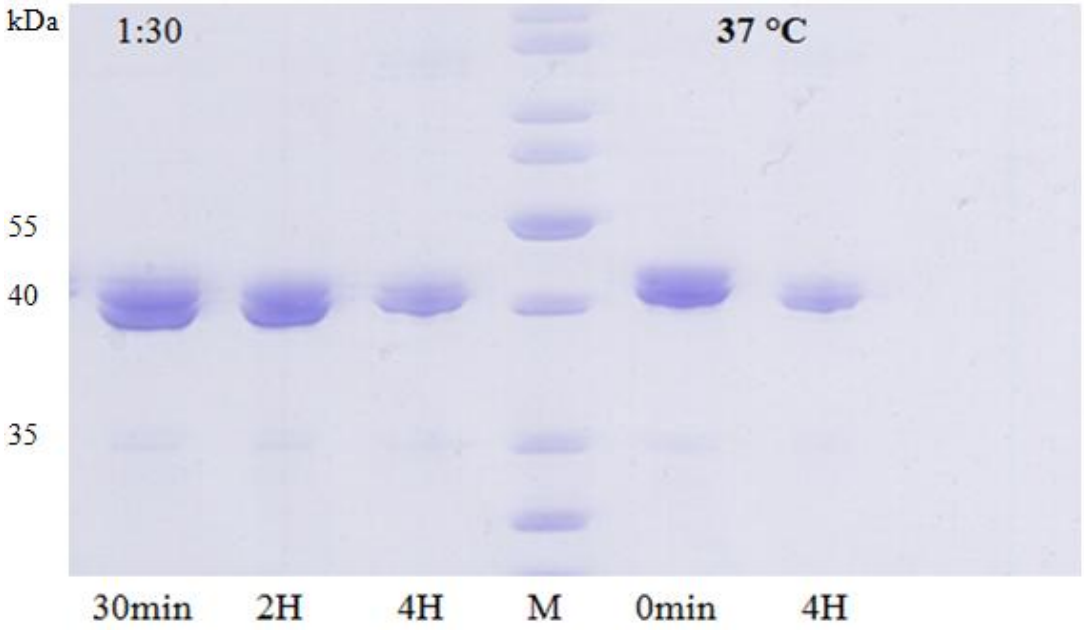


Fig IV.3c Cleavage of tagged SaBIS1 at 37 °C using TEV protease. The SDS-PAGE gel shows efficient cutting of tagged SaBIS1 after 4 h using a 1:30 of TEV protease to SaBIS1. It also shows the slight size differences between the samples before and after removal of the His_6-tag.

After 4 h cleavage, the resulting tag-free SaBIS1 is still active (Figure IV.4).

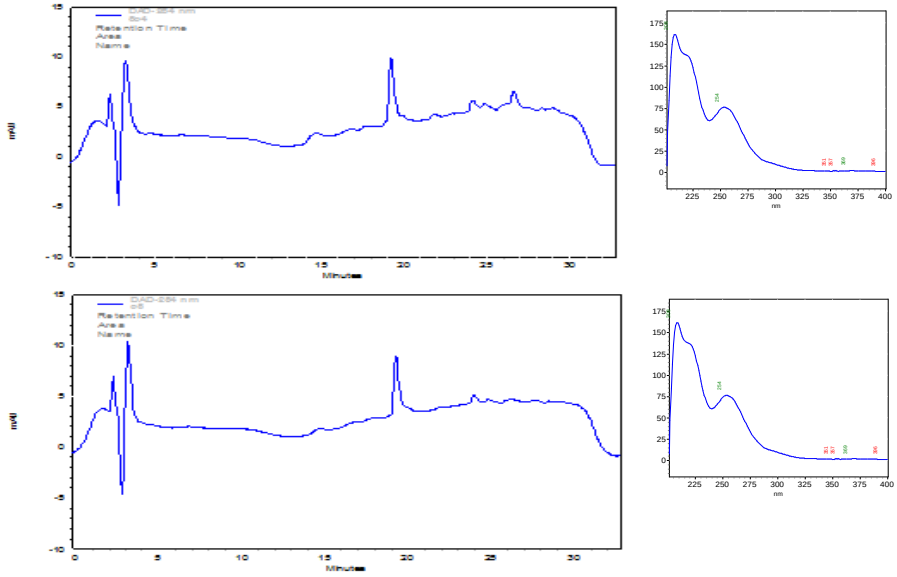


Fig IV.4 Activity of SaBIS1 after 4 h cutting with TEV protease at 37 °C. The peak of the product (3,5-dihydroxybiphenyl) and the corresponding UV spectrum (208 and 254 nm) remained unchanged after cutting.

The cleaved proteins samples were combined and mixed for 1 h with Ni-NTA agarose. The flow through of the column was pooled, concentrated and subjected to gel filtration chromatography using the buffer Tris-HCl pH 7.75 (Lütge, 2012). Although the chromatogram showed a sharp single peak, SDS PAGE of the protein fractions revealed an unexpected and insufficient separation of the cutted and non-cutted enzymes (Figure IV.5).

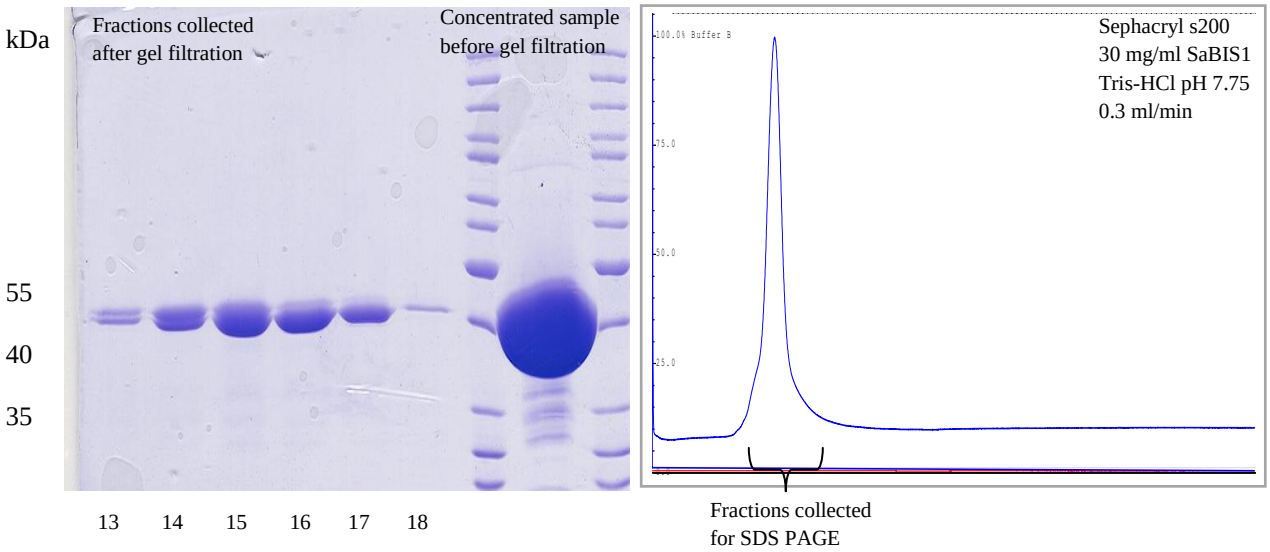


Fig IV.5 Incomplete cutting of N-terminally His_6-tagged SaBIS1. A sharp peak appeared upon gel filtration chromatography, but SDS-PAGE showed a mixture of the cutted and non-cutted enzymes in most protein fractions.

2.2. C-terminally His_6-tagged SaBIS1

Higher concentrations of TEV protease until 1:1 TEV protease to N-terminally His_6-tagged SaBIS1 did not improve the cutting process. The separation of the cutted and non-cutted enzymes was difficult even when we tried to separate them by ion exchange chromatography. Therefore, the new idea was to construct C-terminally His_6-tagged SaBIS1. The presence of the tag at the C-terminus may promote better and eventually complete cutting of tagged SaBIS1 with TEV protease. After primer design (II 3.4.) and PCR (III 1.12.), the new C-terminally His_6-tagged SaBIS1 was produced using the new vector pET-52b (+) (5.2 kb) (Figure IV.6).

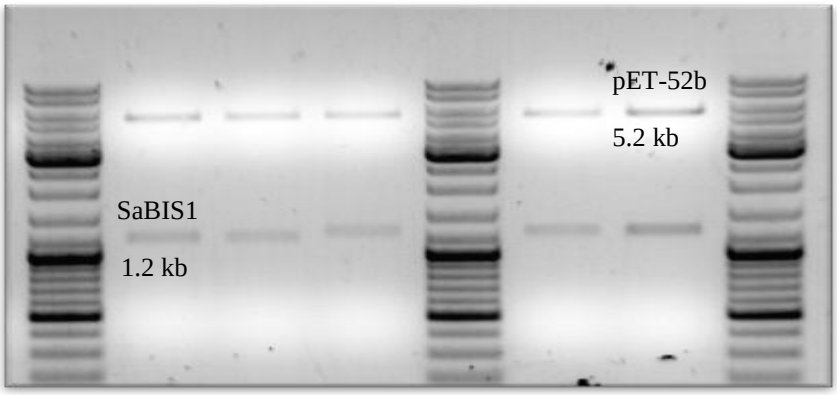


Fig IV.6 Cloning of C-terminally His_6-tagged SaBIS1. The agarose gel presents the double digestion of the constructed plasmid by *Bam*HI and *Sal*I to show the single bands of insert (SaBIS1) and vector (pET-52b).

The C-terminally His_6-tagged SaBIS1 was expressed, purified and cutted using the same conditions previously used for the N-terminally tagged protein (IV 3.1.). The construct exhibited a lower expression level than the one for the N-terminally tagged protein, producing less protein amount as indicated by SDS PAGE. A sharp single peak appeared after gel filtration chromatography, but the fractions again showed a mixture of cutted and non-cutted enzyme proteins (Figure IV.7).

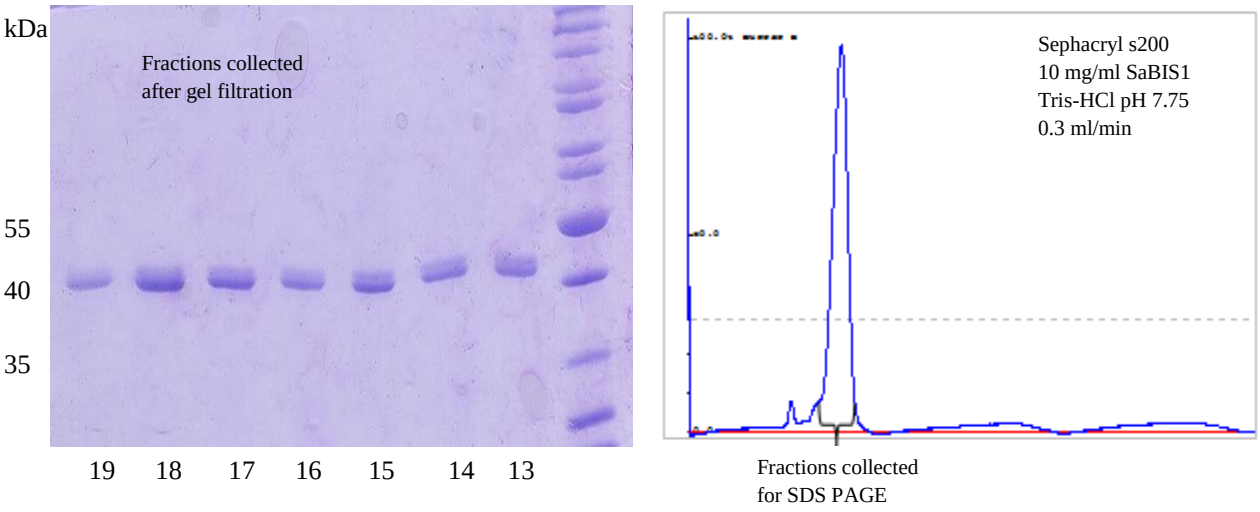


Fig IV.7 Incomplete cutting of C-terminally His_6-tagged SaBIS1. A sharp peak appeared upon gel filtration chromatography, however, SDS PAGE showed a mixture of cutted and non-cutted enzymes in all protein fractions.

Another two truncated versions of SaBIS1 and HaBPS were tested (Lütge, 2012). They have the same amino acids sequences like SaBIS1 and HaBPS, but without the first 13 N-terminal amino acids of SaBIS1 (BIS-13AA) and without the first 19 N-terminal amino acids of HaBPS (BPS-19AA). The removed amino acids represent disordered regions in the beginning of the sequences, analogous to the already crystallized MsCHS2. Firstly, both enzymes were tested to ensure their activity after the shortening process (Figure IV.8).

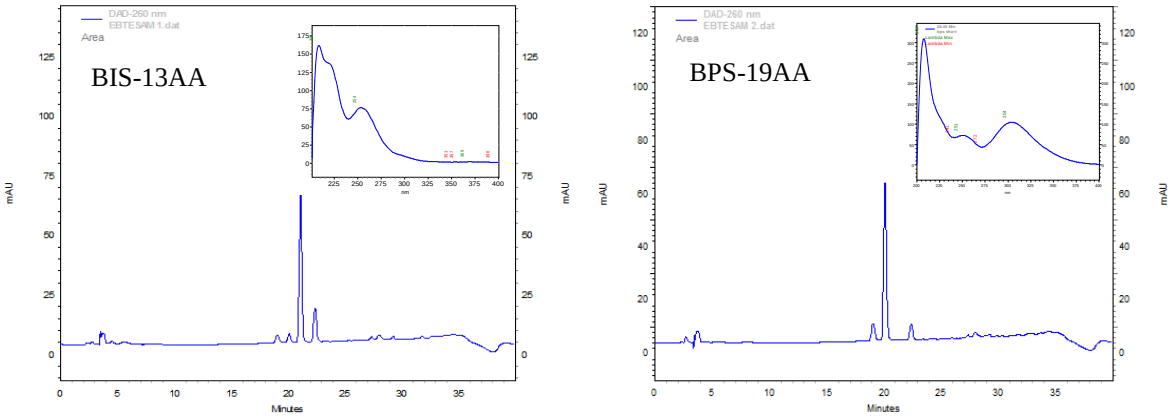


Fig IV.8 Activities of BIS-13AA and BPS-19AA. HPLC profiles and corresponding UV spectra of the BIS product (3,5-dihydroxybiphenyl; 208 and 254 nm) and the BPS product (2,4,6-trihydroxybenzophenone; 208 and 304 nm) are shown.

Unfortunately, both purified and cutted proteins, BIS-13AA and BPS-19AA, showed sharp single peaks after gel filtration chromatography but again mixtures of the cutted and non-cutted enzymes in SDS-PAGE (Figure IV.9).

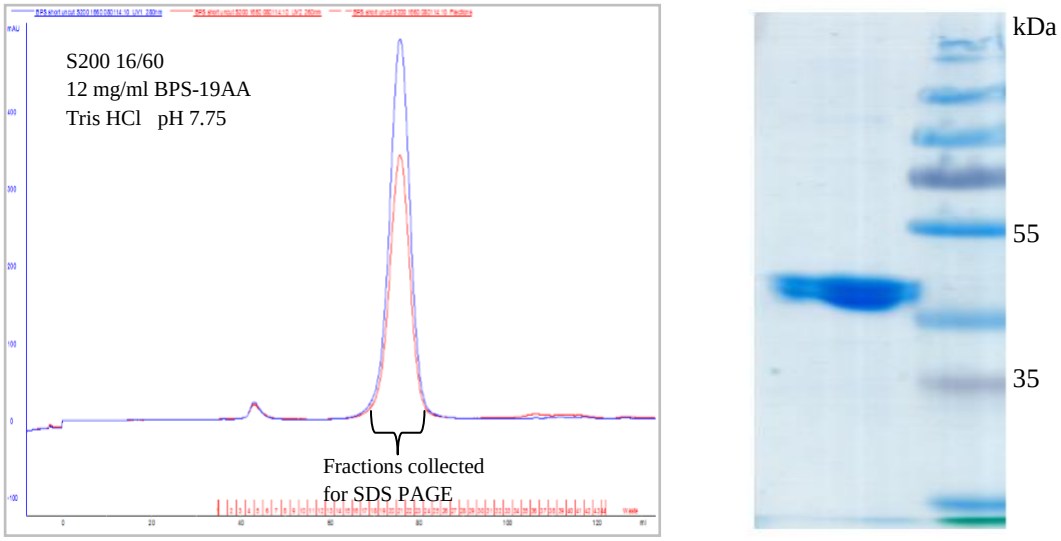


Fig IV.9 Purification and cleavage of tagged BPS-19AA. The SDS-PAGE gel shows the incomplete cutting process and the appearance of a double band despite a sharp peak in the gel filtration chromatogram.

The cutting process with all the previously used constructs and conditions was always unsuccessful. In addition, the two-step purification, before and after cutting the enzymes, is time-consuming, not to mention that the cutting process itself takes time. Also, the extensive purification steps due to the cutting process led to a significant decrease in the final protein amount, which in turn is a problem for the crystallization process, as we should end up with a highly concentrated protein sample. All these above-mentioned difficulties led us to think about purification and crystallization of the full His_6-tagged protein without any cutting. It was meanwhile proved that the purification tags in many proteins had no significant effect on the structures of the native proteins (Carson et al., 2007). Therefore, the work was started from the beginning, using the N-terminally His_6-tagged constructs of both enzymes (MdBIS and HaBPS). In addition, optimization was carried out to improve the purification protocol and to select buffers for ensuring a good quality of highly purified proteins for crystallization and structure elucidation.

3. Expression and purification of large amounts of target proteins

The first requirements for protein crystallization are **high purity** (≥ 98%, as indicated by a single band on a SDS-PAGE gel) and **high concentration** (≥ 5-10 mg/ml). Therefore, a good expression and purification system was developed.

3.1.Affinity chromatography and PD-10

All the enzyme proteins used in this work were cloned and expressed with a N-terminal His_6-tag in the expression vector pRsetB, enabling purification from *E. coli* using Ni-NTA-agarose. Freshly transformed bacteria and new plates were prepared before any crystallization process of target proteins. Large amounts of proteins had to be produced and therefore large culture volumes were used. The individual proteins expressed differently and hence different culture volumes were used to get the desired concentration, for example, SaBIS1 needed a larger culture volume to get highly concentrated protein in comparison to the other enzymes. The pellets were harvested and subjected to affinity chromatography (III.2.1.1). Optimization of the purification was carried out using buffers containing different concentrations of imidazole (30 mM, 50 mM and 80 mM) to ensure the elimination of all foreign proteins. For larger pellets, as with SaBIS1, the column was washed extensively by increasing the washing times and steps. The eluates were collected and buffer exchange was carried out using PD-10 (III.2.1.2.1.). The pooled fractions were concentrated using a Vivaspin concentrator (III.2.2.) and the protein content was determined using the Bradford method (III.2.3.). Finally, the purity was checked using SDS-PAGE (III.2.5.) (Figure IV.10).

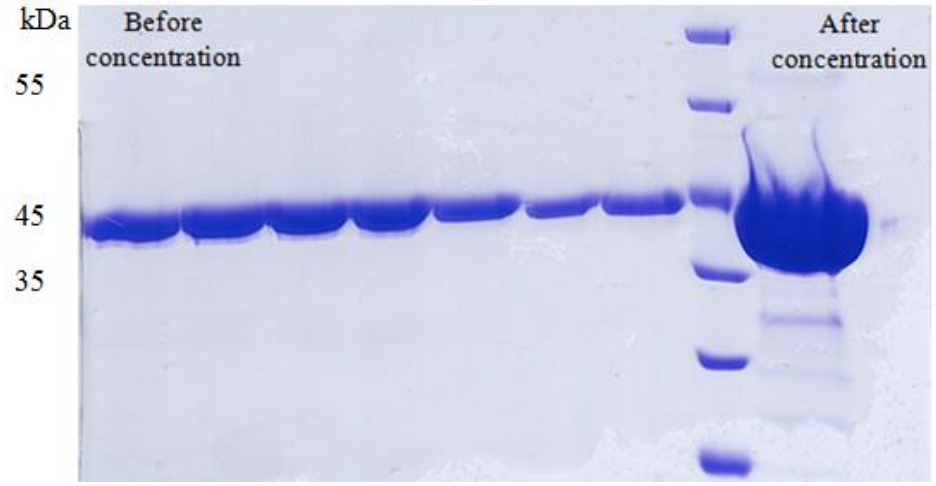


Fig IV.10 SDS-PAGE of purified SaBIS1. The gel shows the purified fractions before concentration and the final concentrated sample of purified protein.

The differences in the purification steps and culture volumes of all target proteins are illustrated in the following Table 1:

Table 1. Survey of the expression and purification conditions used for the various target proteins

Protein	SaBIS1 HsBPST135K	MdBIS3, MdBIS3C159A	HaBPS,HaBPSC167A HaBPST135L
Culture volume	5-7 L	3-5 L	3 L
Produced amount	20-40 mg	20-30 mg	20 mg
IPTG concentration	0.125 Mm	0.1875 Mm	0.25 mM
Washing steps: 30 mM imidazolebuffer 50 mM imidazole buffer 80 mM imidazole buffer	 16 times (1 ml) 10 times (1 ml) 8 times (1 ml)	 12 times (1 ml) 8 times (1 ml) 6 times (1 ml)	 10 times (1 ml) 6 times (1 ml) 4 times (1 ml)

The time for culturing or bacteria growing and induction was optimized to perform the purification process in minimum time intervals and to keep the protein as fresh as possible for crystallization. The first day was for preculturing. The second day was for culturing for 3-4 h until OD 1, induction for 4 h and pellet storage at -20 °C. The third day at TUBS was for protein purification, concentration, SDS-PAGE and carrying the protein sample to HZI, where, the sample was directly subjected to gel filtration. In the fourth day, the protein fractions were concentrated, another SDS-PAGE was performed and finally the crystallization screens were done.

3.2. Thermofluor assay and buffer optimization

The second important condition for the crystallization is the **stability** of the protein, which was targeted by screening the best buffer conditions for the enzyme using the thermofluor assay. This test was conducted in cooperation with Dr. Peer Lukat (HZI, Braunschweig). The following six buffers were used to test the stability of the protein: acetate, citrate, MES, HEPES, tris-HCl and CHES. They were tested at different pH values and salt (NaCl) concentrations. Thereafter, different additives like glycerol, DTT, PEG400 and galactose were tested for their effect on the stability of the protein in the respective buffer (III 4.4.). High melting temperatures between 47-51 C° were recorded for the best buffer conditions for protein stability. Such values were recorded for SaBIS1 after using the buffers 50 mM Tris-HCl and 50 mM HEPES. After further screens for additives, salts and pH, the following buffer was selected as the buffer of choice for SaBIS1 (Figure IV.11a, b):

50 mM Tris-HCl, 100 mM NaCl, 10% glycerol, pH 7.4

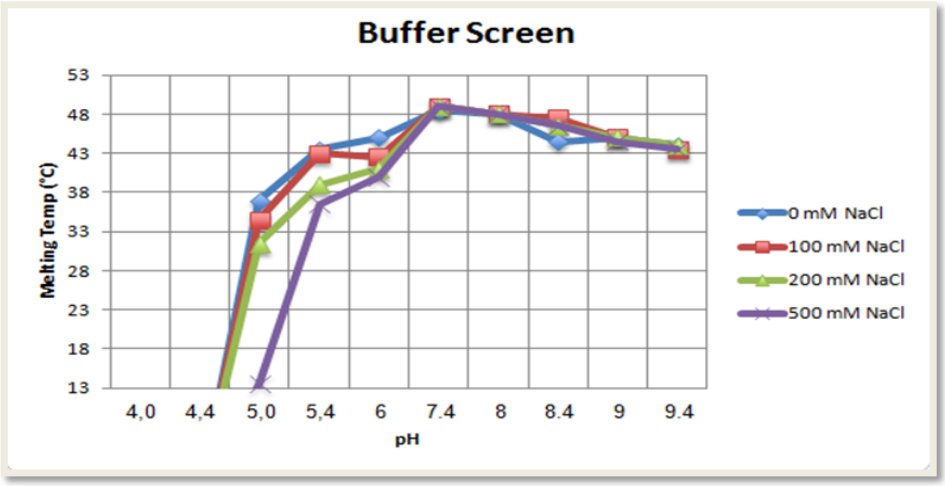


Fig IV.11a Thermofluor assay results to detect the best buffer for SaBIS1. A level of 100 mM NaCl resulted in high stability of the protein in Tris-HCl buffer pH 7.4.

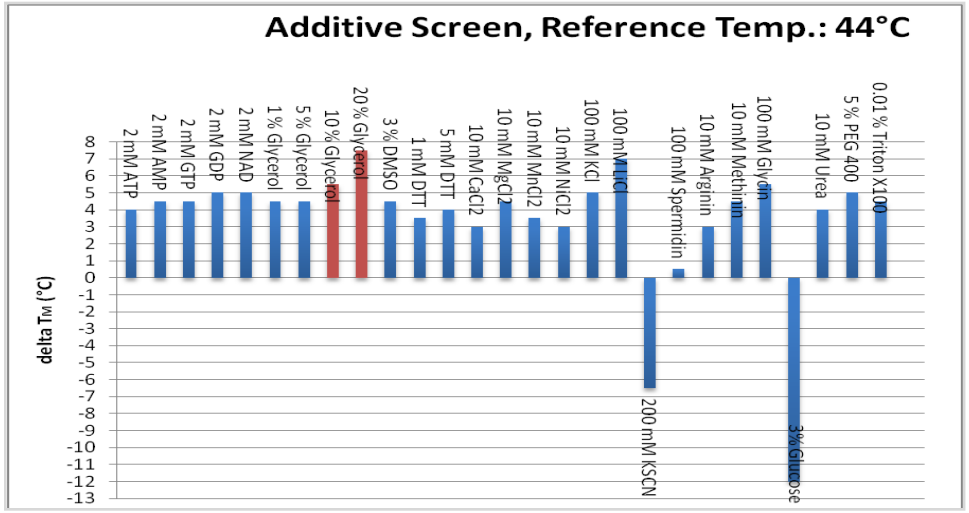


Fig IV.11b Thermoflour assay results to detect the best additive for SaBIS1. Concentrations of 10% and 20% glycerol (red) showed high thermal signals, which indicated better stability of SaBIS1 in this buffer.

Similar melting temperatures were recorded for MdBIS3 using the buffer 50mM Tris-HCl. After further screens for additives, salts and pH, the following buffer turned out to be the choice for MdBIS3 (Figure IV.12a, b):

50 mM Tris-HCl, 300 mM NaCl, 10% glycerol, 1% galactose, pH 8.4

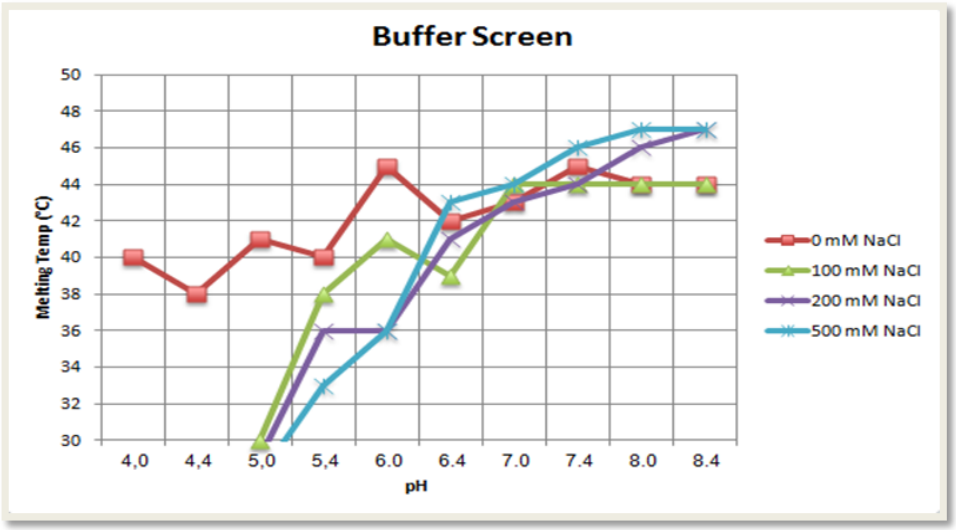


Fig IV.12a Thermoflour assay results to detect the best buffer for MdBIS3. Levels of 200-500 mM NaCl resulted in high stability of the protein in Tris-HCl buffer pH 8.4.

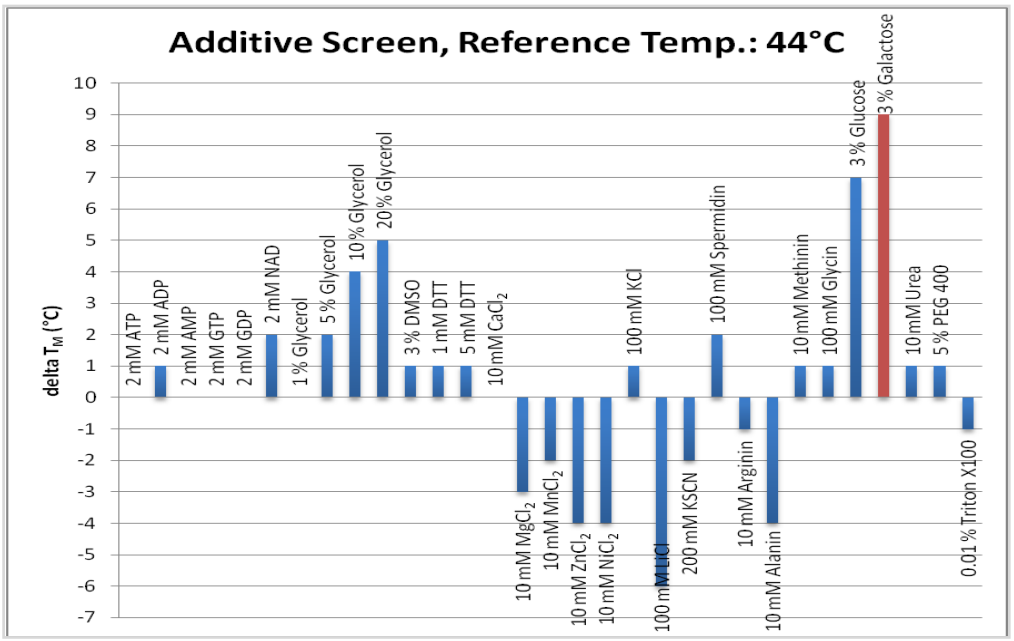


Fig IV.12b Thermoflour assay results to detect the best additive for MdBIS3. A concentration of 3% galactose (red) showed a high thermal signal, which indicated better stability of MdBIS3 in this buffer.

In case of HaBPS, similar values were recorded as for SaBIS1, also resulting in the same buffer (50 mM Tris-HCl). After further screens for additives, salts and pH, the following buffer was picked as the choice for *Hypericum* BPSs (Figure IV.13a, b):

50 mM Tris-HCl, 100 mM NaCl, 10% glycerol, pH 8.4

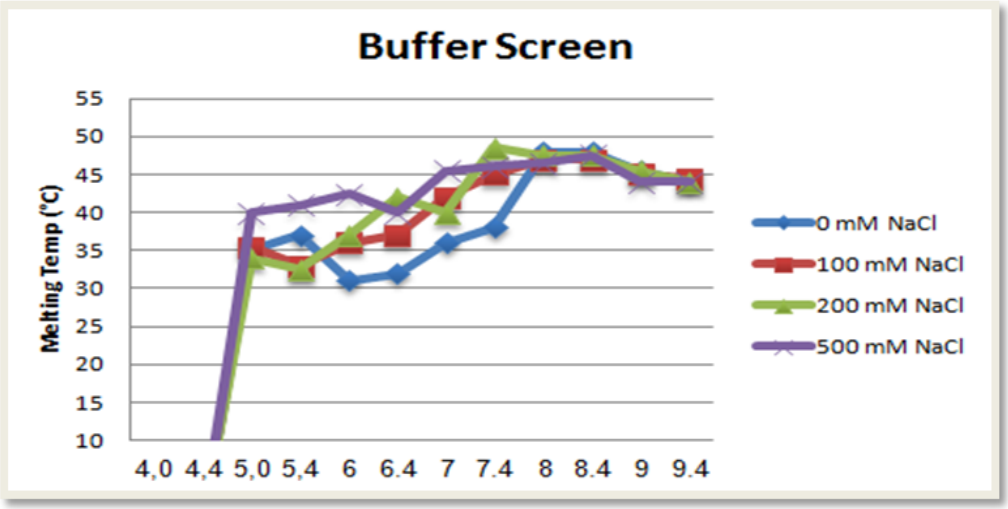


Fig IV.13a Thermoflour assay results to detect the best buffer for HaBPS. A level of 100 mM NaCl, as the other higher concentrations, resulted in high stability of HaBPS in Tris-HCl buffer pH 8.4.

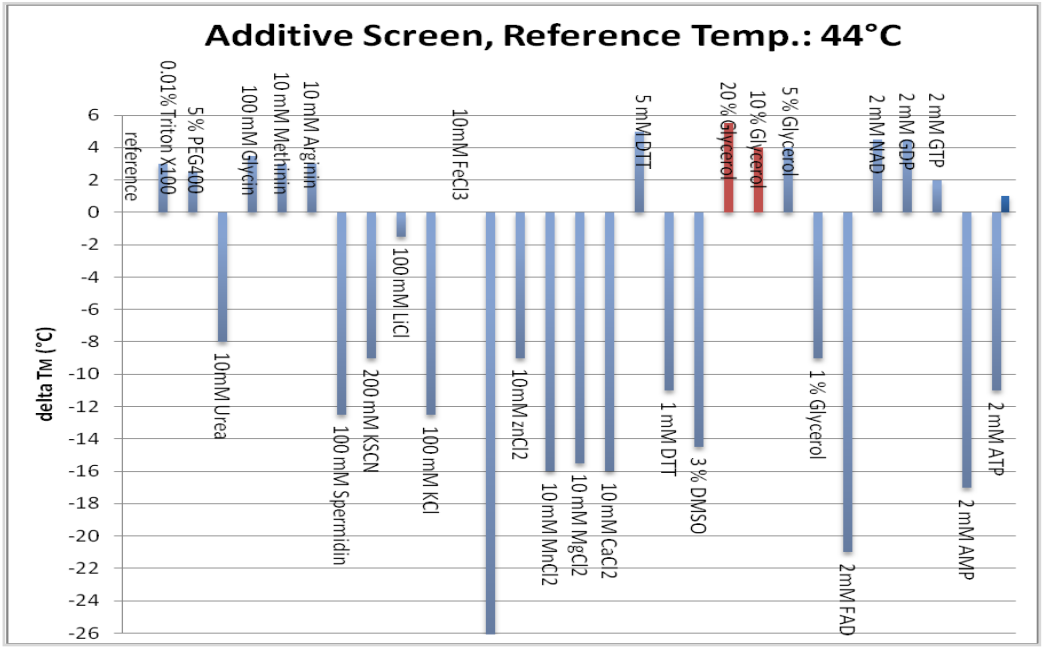


Fig IV.13b Thermoflour assay results to detect the best additive for HaBPS. Concentrations of 10% and 20% glycerol (red) showed high thermal signals, which indicated better stability of HaBPS in this buffer.

3.3. Gel filtration and ion exchange chromatography

The third important condition for the crystallization is protein **homogeneity**, which is indicated by a single sharp peak of the dimer in the chromatogram of gel filtration on sephacryl and superdex columns (III 2.1.2.2). This chromatography was conducted in cooperation with Dr. Peer Lukat (HZI, Braunschweig). The purpose of the separation was to get rid of both any aggregates that had formed during the purification process and any other impurities that were difficult to remove by affinity chromatography. For SaBIS1, there were always four main peaks in the chromatogram (Figure IV.14). It was essential to separate them first by ion exchange chromatography using a Mono-Q column (III 2.1.3.1). In the end, only the single peak of the dimer should be left.

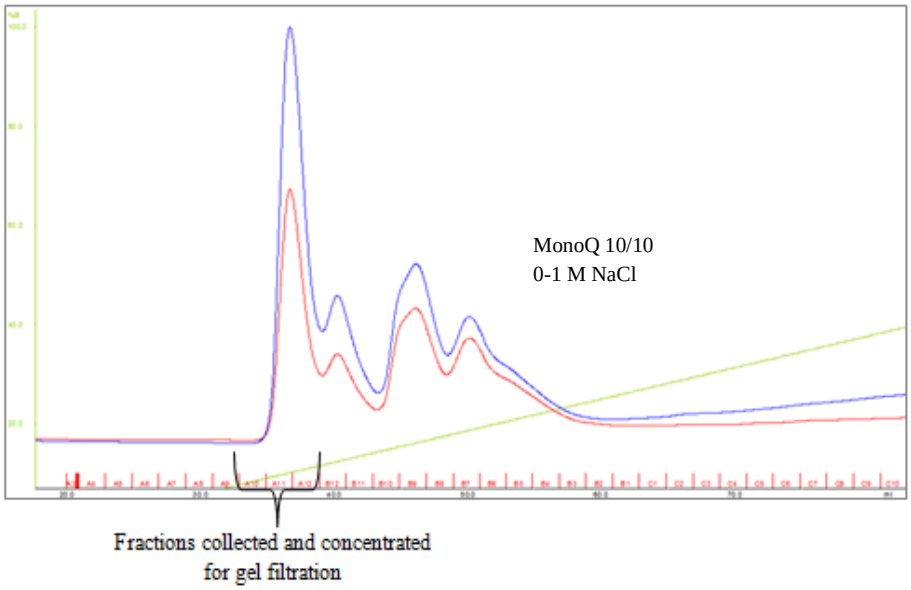


Fig IV.14 Ion exchange chromatography of SaBIS1. Four major peaks were separated.

After ion exchange chromatography, the fraction of the dimer, i.e. the first fraction, was collected and subjected to gel filtration chromatography to confirm single peak formation (Figure IV.15).

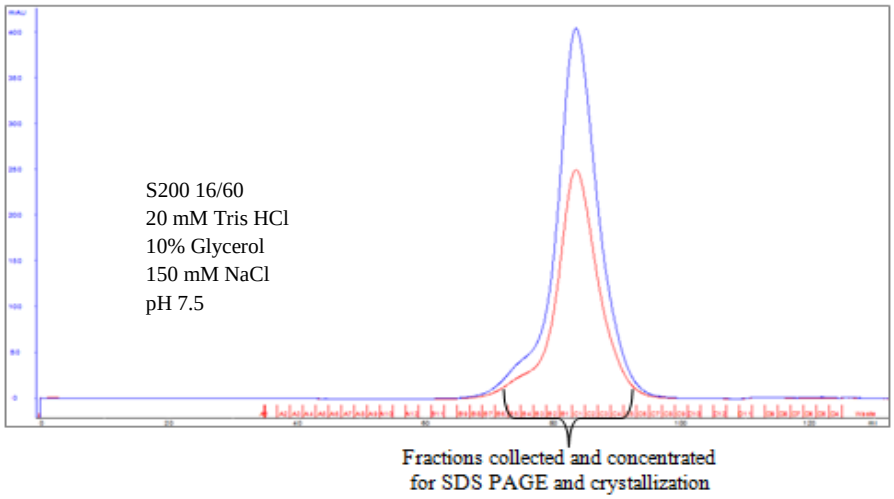


Fig IV.15 Single peak formation by SaBIS1 after ion exchange and gel filtration chromatographies.

The homogeneity of MdBIS3 and Hypericum BPSs was indicated by single sharp peaks of the dimers in the chromatograms of superdex columns (III 2.1.2.2.), which were here sufficient for purification (Figure IV.16).

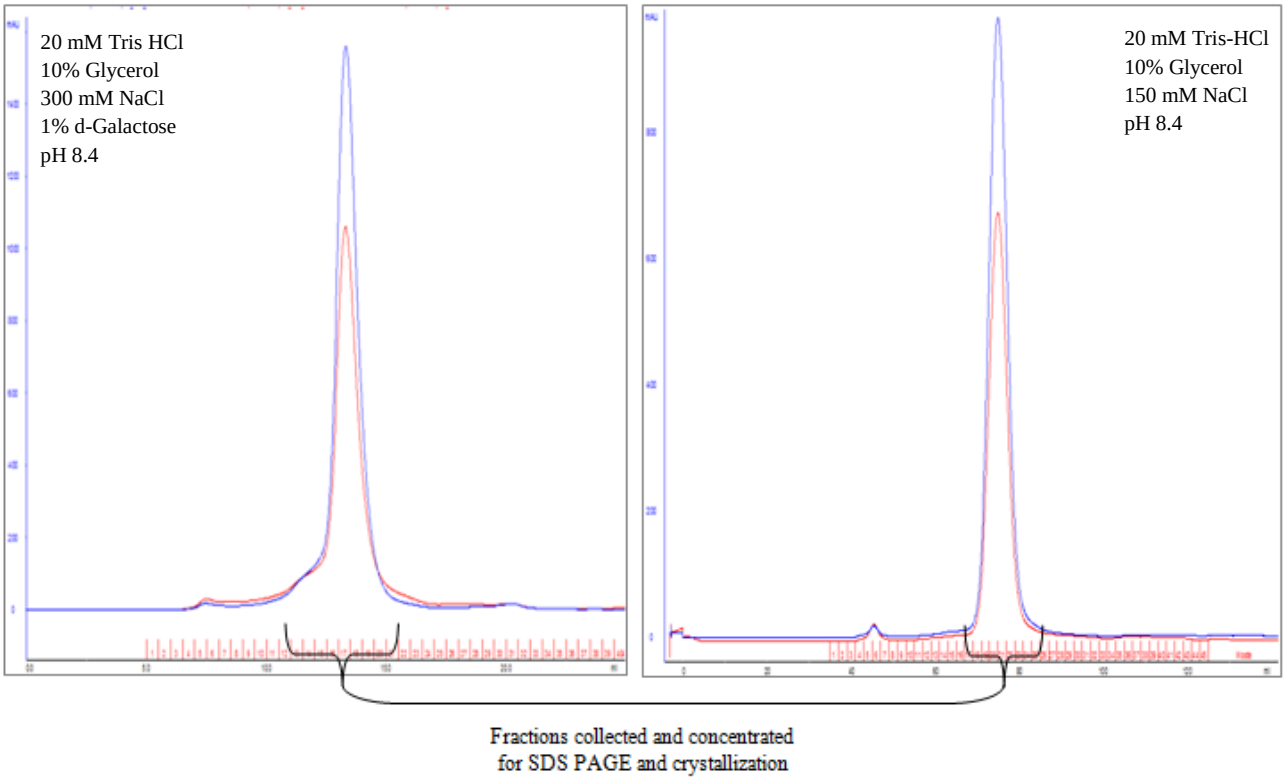


Fig IV.16 Single peak formation by MdBIS3 (left) and HaBPS (right) in gel filtration chromatography. Additional ion exchange chromatography was not necessary.

3.4. Final SDS-PAGE and purity confirmation

A number of purification processes like affinity, ion exchange, and gel filtration chromatographies were performed to purify the proteins. In addition, many different buffers like washing, elution, and desalting buffers containing different components were used. The proteins were at each step in contact with different compounds and colums, therefore, it was essential to conduct another SDS-PAGE after the last purification step and just before the crystallization process to ensure the final purity of the target protein and to eliminate any risk of impurities that could have been unintentionally formed (Figure IV.17).

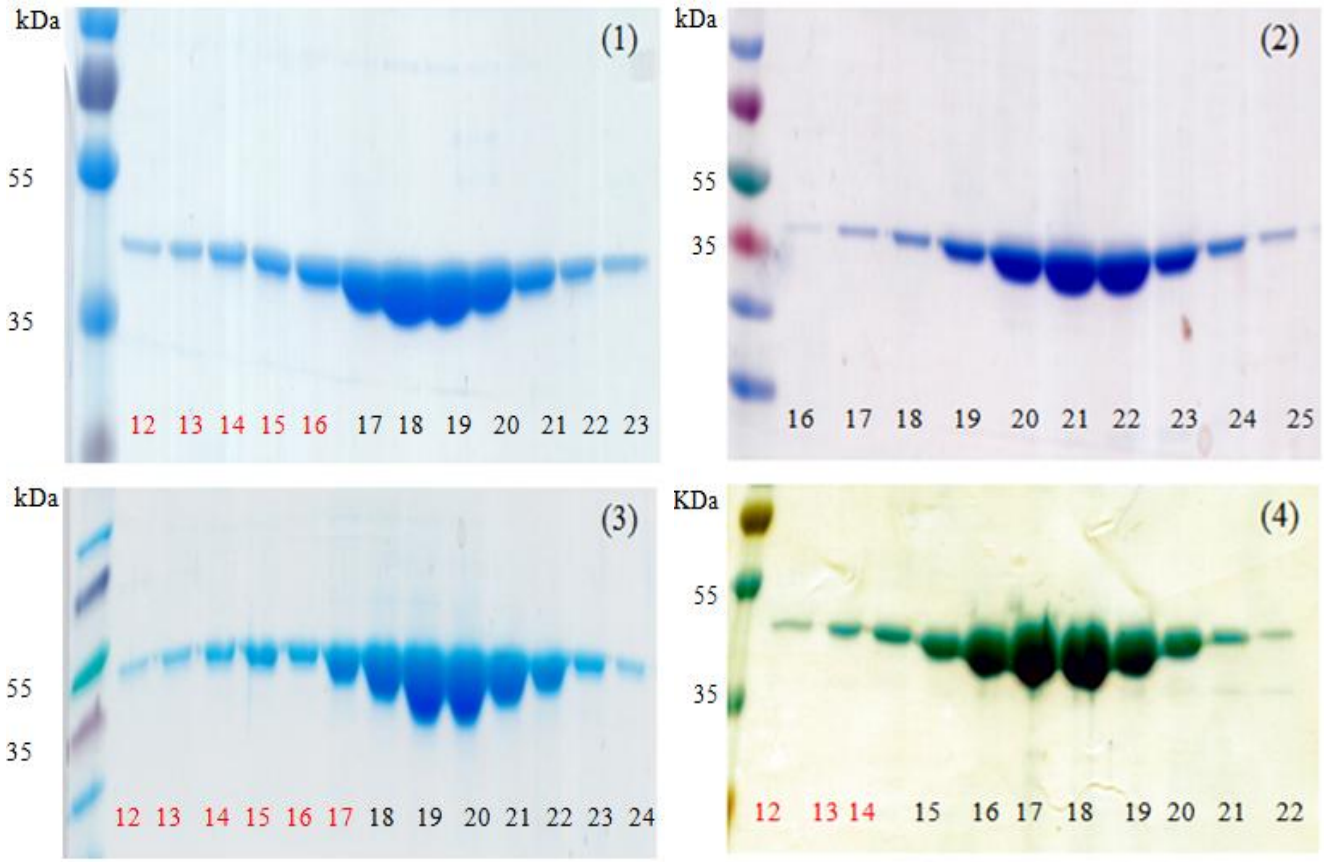


Fig IV.17 SDS-PAGE for final purity confirmation, before crystallization, of 1) HaBPS, 2) MdBIS3, 3) HaBPST135L and 4) HsBPS. Black numbers represent the protein fractions after gel filtration chromatography, which corresponded to the main sharp peak of the protein. Red numbers correspond to the non-specific hump at the beginning of the sharp peak.

3.5. Estimation of Enzymes activities

Since the enzymes were subjected to many different purification steps, it was essential to estimate their catalytic activities with their preferred substrates (III 3.1.) before going through the crystallization process. The formation of 3,5 dihydoxy biphenyl was confirmed after reaction of SaBIS1 and MdBIS3 with benzoyl-CoA and malonyl-CoA (Figure IV.18a). MdBIS3 was also tested for its activity with salicoyl-CoA and the formation of 4-hydroxycoumarin was confirmed (Figure IV.18b).

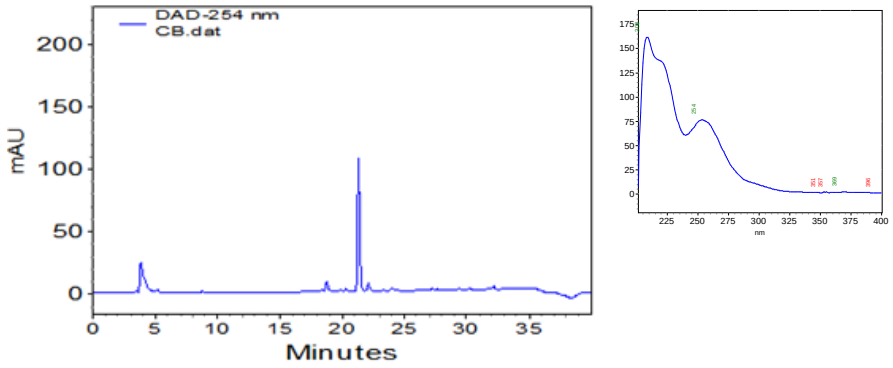


Fig IV.18a HPLC analysis of the formation of 3,5-dihydroxybiphenyl by SaBIS1. The UV spectrum of the product matched that of an authentic reference compound with absorption maxima at 208 and 254 nm.

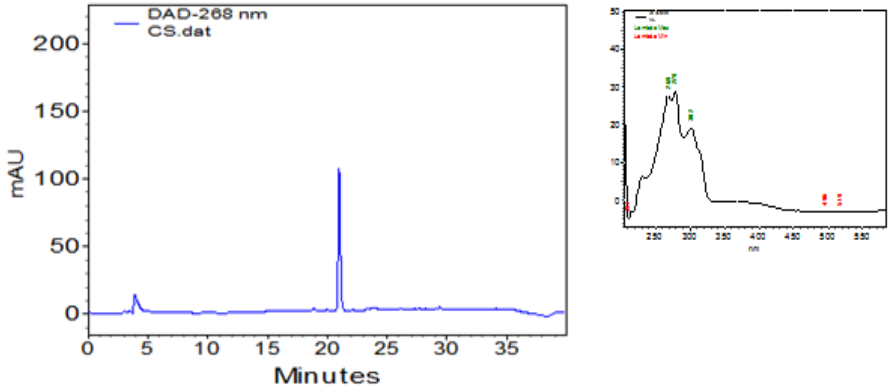


Fig IV.18b HPLC analysis of the formation of 4-hydroxycoumarin by MdBIS3. The UV spectrum of the product matched that of an authentic reference compound with absorption maxima at 260, 265 and 305 nm.

The activities of HaBPS and HsBPS with benzoyl-CoA and malonyl-CoA were confirmed by formation of 2,4,6-trihydroxybenzophenone (Figure IV.19).

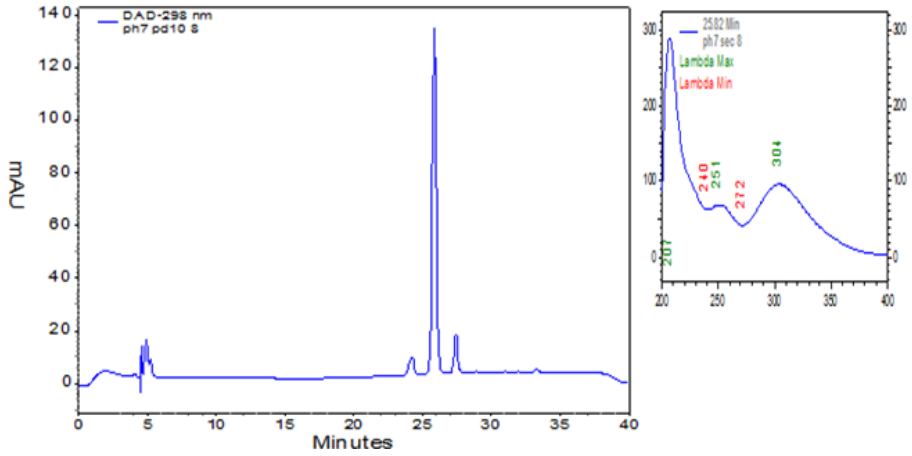


Fig IV.19 HPLC analysis of the formation of 2,4,6-trihydroxybenzophenone by BPS. The UV spectrum of the product matched that of an authentic reference compound with absorption maxima at 202, 251 and 304 nm.

3.6. Purification at acidic pH

According to the thermofluor test, a pH value of 8.4 was mainly used in the purification of the enzymes. This pH did not affect the activities of BPS and BIS, as shown by HPLC analysis of enzyme assays (IV 4.2.), although they are more active at slight acidic conditions (pH 5-7). The stability of the CoA substrates may be affected after soaking and co-crystallizaion processes, as the basic pH may lead to substrate degredation. Therefore, the possibility of using a slightly acidic pH was tested to see its effect on the purification process, the stability of the protein and the peak shape after gel filtration chromatography (Figure IV.20).

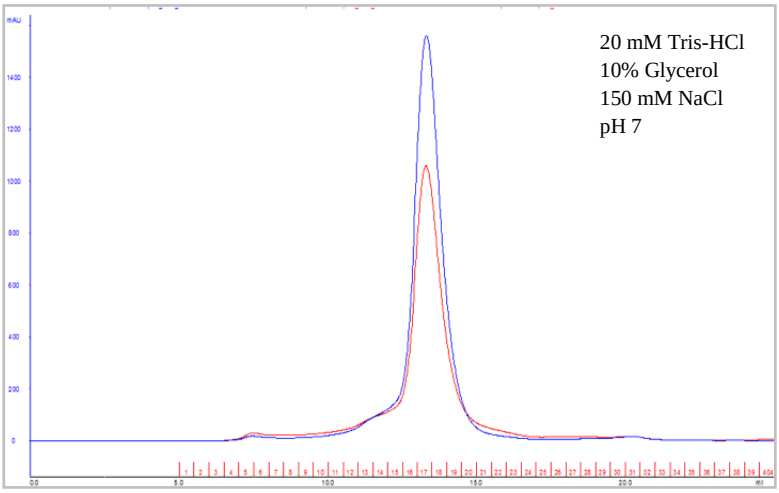


Fig IV.20 Purification of HaBPS at slightly acidic pH. At pH 7, HaBPS showed a sharp peak after purification by gel filtration chromatography using Tris-HCl buffer.

Further decreasing the pH below 7 was also tested for the purification of HsBPS. The acidic pH (pH 6.7) resulted in distortion of the peak shape after gel filtration, which means that BPSs are not stable at acidic conditions and show aggregation (Figure IV.21). The enzymes prefer a more basic pH and therefore we maintained the purification conditions at the previously used basic conditions.

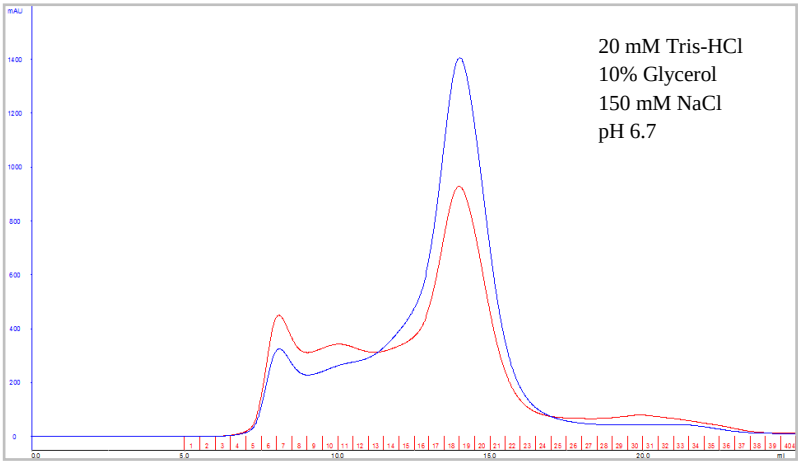


Fig IV.21 Purification of HsBPS at slightly acidic pH. At pH 6.7, HsBPS showed an unorganized disturbed peak, which may indicates the unstability and aggregation at acidic conditions.

4. Crystallization and buffer screens

4.1. *Malus domestica* BIS3

Crystallization of PKSs was carried out in cooperation with Dr. P. Lukat and Prof. W. Blankenfeldt, HZI, Braunschweig. The first aim was the crystallization of SaBIS1. After purification optimization (IV.4), pure protein of high quality for the crystallization process was obtained. However, the use of various commercial buffer screens (JCSG+ 1-4, Morpheus, Index and Midas) failed to yield crystal formation; only precipitate was produced under these conditions. Here the idea of trying another BIS, namely MdBIS3, emerged. The protein solutions used had concentrations of 10 mg/ml and 20 mg/ml. Drops (0.2 µl protein solution and 0.2 µl well content) were placed on a plate containing 60 µl well volume using a robot. The plates were sealed and kept at 25 °C under continuous visualization using an imager. The following Table 2 shows the best buffer screens tested for MdBIS3, which developed good plate crystals yielding a satisfactory diffraction pattern after X-ray measurement (Figure IV.22).

Table 2. Buffer screen for MdBIS3

Crystallization kit	Composition		Fig IV.22 panel No.
Morpheus screen A12 No 12	Morpheus buffer 3 (0.1 M) pH 8.5 Morpheus Divalents (0.06 M) MPD_P1k_P3350 (37.5% w/v) Water	6 µl 6 µl 30 µl 18 µl	1
Morpheus screen E12 No 60	Morpheus buffer 3 (0.1 M) pH 8.5 Morpheus Ethylene glycols (0.12 M) MPD_P1k_P3350 (37.5% w/v) Water	6 µl 6 µl 30 µl 18 µl	2
Morpheus screen F12 No 72	Morpheus buffer 3 (0.1 M) pH 8.5 Morpheus monosaccharides (0.12 M) MPD_P1k_P3350 (37.5% w/v) Water	6 µl 6 µl 30 µl 18 µl	3
JCSG Core III B5 No 17	Tris (0.1 M) pH 8.5 Sodium acetate (0.2 M) Polyethylene glycol 4000 (30 % w/v) Water	6 µl 4 µl 36 µl 14 µl	4
JCSG Core IV E12 No 60	HEPES (0.1 M) pH 6.5 Polyethylene glycol 6000 (30 % w/v) Water	6 µl 36 µl 18 µl	5
Index screen G9 No 81	Tris (0.1 M) pH 8.5 Amonium acetate (0.2 M) Polyethylene glycol 3.350 (25% w/v) Water	6 µl 12 µl 30 µl 12 µl	6

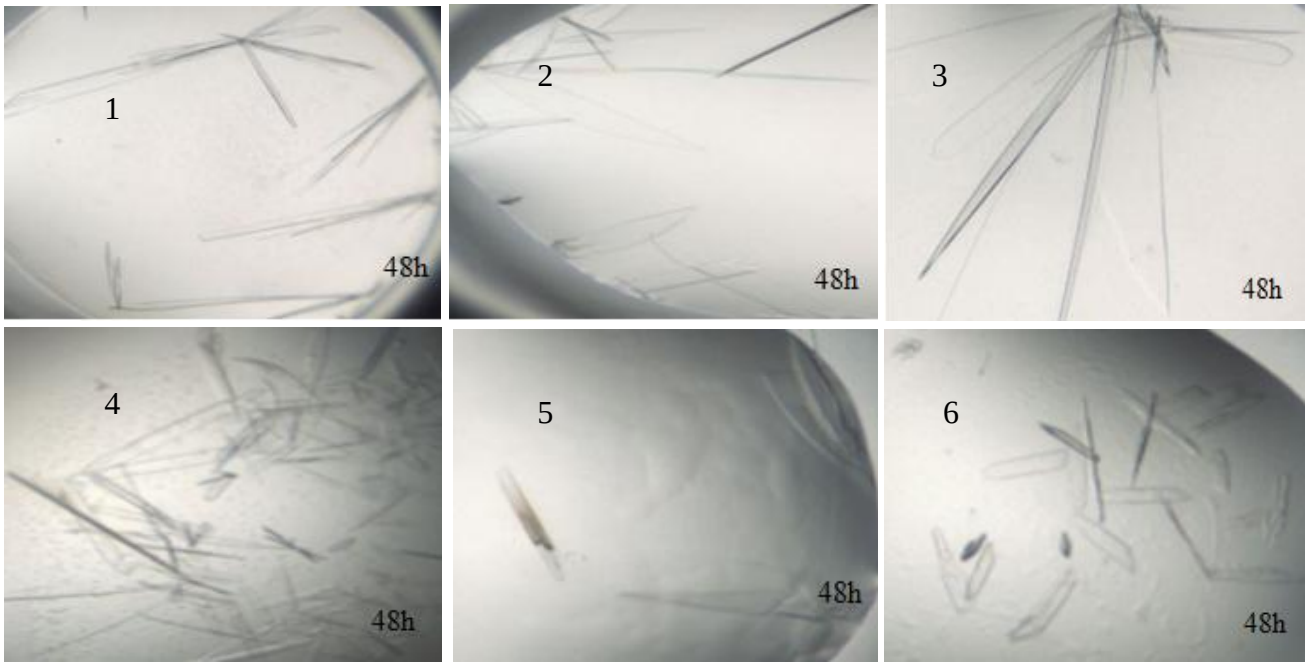


Fig IV.22 Crystallization hits for MdBIS3. The numbers indicate the different buffers screens, as listed in Table 2. All crystals developed after 48 h.

4.2. *Hypericum androsemum* BPS and BPST135L

The protein solution used had concentrations of 2, 4, 5, 8 and 10 mg/ml. The following Table 3 presents the different buffers screens, under which needle-shaped crystals developed (Figure IV.23).

Table 3. First buffer screen for HaBPS

Crystallization kits	Composition		Fig IV.23 panel No.
Random II screen C9 No 33	MES (0.1 M) pH 6.233 D-Sorbitol (0.133 M) Polyethylene glycol 4000 (24.4% w/v) Water	12 µl 8 µl 29.333 µl 10.667 µl	1
Random II screen F3 No 63	HEPES (0.1 M) pH 7 D-Sorbitol (0.106 M) Polyethylene glycol 4000 (25% w/v) Glycerol (13.9% w/v) Water	6 µl 6.333 µl 37.5 µl 9.69 µl 0.477 µl	2
Random II screen G3 No 75	HEPES (0.1 M) pH 6.9333 Polyethylene glycol 200 (10% w/v) Polyethylene glycol 8000 (25% w/v) Water	6 µl 6 µl 37.5 µl 10.5 µl	3
Random II screen G10 No 82	MES (0.1 M) pH 6.311 Polyethylene glycol 4000 (21.1% w/v) Water	12 µl 25.333 µl 22.667 µl	4

Crystallizations-Kits	Composition		Fig IV.23 panel No.
Random I screen A4 No 4	MES (0.1 M) pH 6.467 Glycerol (9.44% w/v) Polyethylene glycol 8000 (26.7% w/v) Water	12 µl 6.589 µl 40 µl 1.411 µl	5
Random I screen A11 No 11	HEPES (0.1 M) pH 6.889 Polyethylene glycol 200 (11.7% w/v) Polyethylene glycol 8000 (21.7% w/v) Water	6 µl 7 µl 32.5 µl 14.5 µl	6
Random I screen C6 No 30	HEPES (0.1 M) pH 6.933 Ammonium acetate (0.16 M) Polyethylene glycol 200 (11.7% w/v) Polyethylene glycol 3000 (20% w/v) Water	6 µl 4.833 µl 7 µl 24 µl 18.167 µl	7
Random I screen D11 No 47	HEPES (0.1 M) pH 6.882 Glycerol (22.8% w/v) Polyethylene glycol 6000 (26.7% w/v) Water	6 µl 15.891 µl 32 µl 6.109 µl	8
Random I screen G3 No 75	MES (0.1 M) pH 6.233 Polyethylene glycol 6000 (26.7% w/v) Water	12 µl 32 µl 16 µl	9

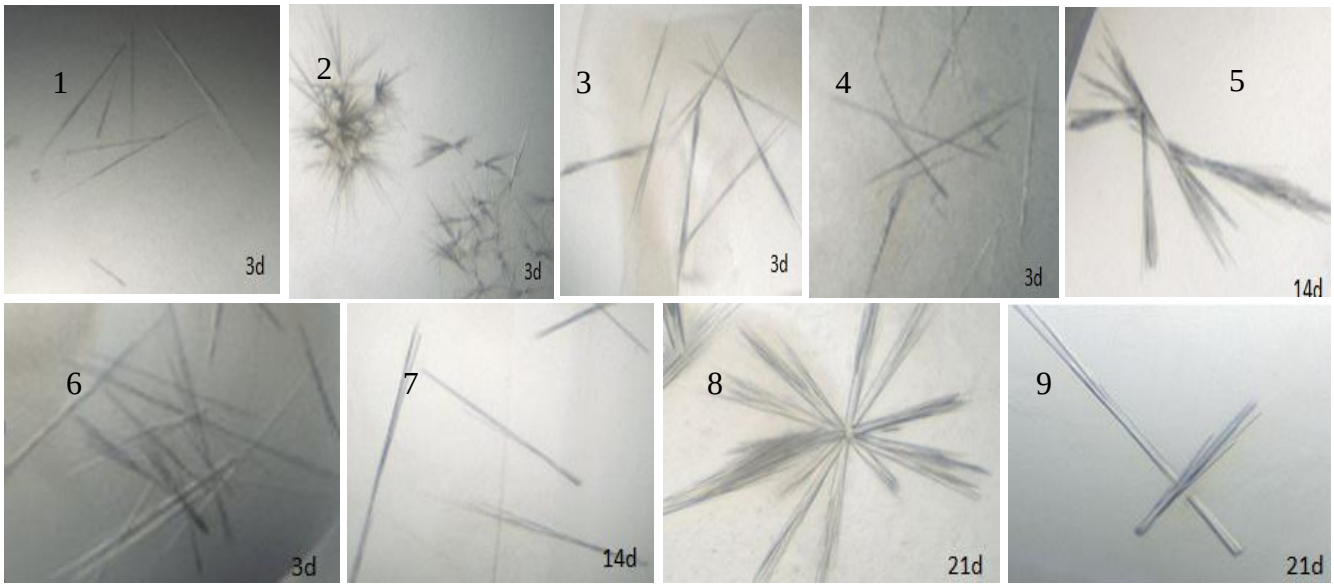


Fig IV.23 Crystallizations hits for HaBPS. The numbers refer to the corresponding conditions in Table 3. The needle-shaped crystals developed after 3 d and continued to grow until 21 d.

Bigger crystals with plate shape developed for HaBPS and HaBPST135L after further optimization and using higher protein concentrations (15 mg/ml and 28 mg/ml) (Figure IV.24). The optimization mainly depended on the use of the specific salt Poly (acrylic acid sodium salt) 2100 in the buffer screens. The conditions are shown in the next Table 4.

Table 4. Second buffer screen for HaBPS and mutants

Enzymes	Crystallization kit	Composition		Fig IV.24 panel No.
HaBPS	Midas screen E7 No 55	HEPES (0.1 M) pH 7 di-Sodium malonate 0.1 M Poly (acrylic acid sodium salt) 2100 (30% w/v) Water	6 µl 2 µl 36 µl 16 µl	1
	Midas screen E11 No 59	HEPES (0.1M) pH 6.5 Lithium sulfate 0.1 M Poly (acrylic acid sodium salt) 2100 (25% w/v) Water	6 µl 3 µl 30 µl 21 µl	2
T135L	BPS Polyacrylate screen A7 No 7	Tris-HCl (0.1M) pH 7.889 Lithium sulfate 0.217 M Poly (acrylic acid sodium salt) 2100 (26.7% w/v) Water	6 µl 6.5 µl 32 µl 15.5 µl	3
	BPS Polyacrylate screen G7 No 78	Tris-HCl (0.1M) pH 7.444 Ammonium sulfate 0.0778 M Glycerol (7.22% w/v) Poly (acrylic acid sodium salt) 2100 (25% w/v) Water	6 µl 1.333 µl 5.039 µl 30 µl 17.628 µl	4

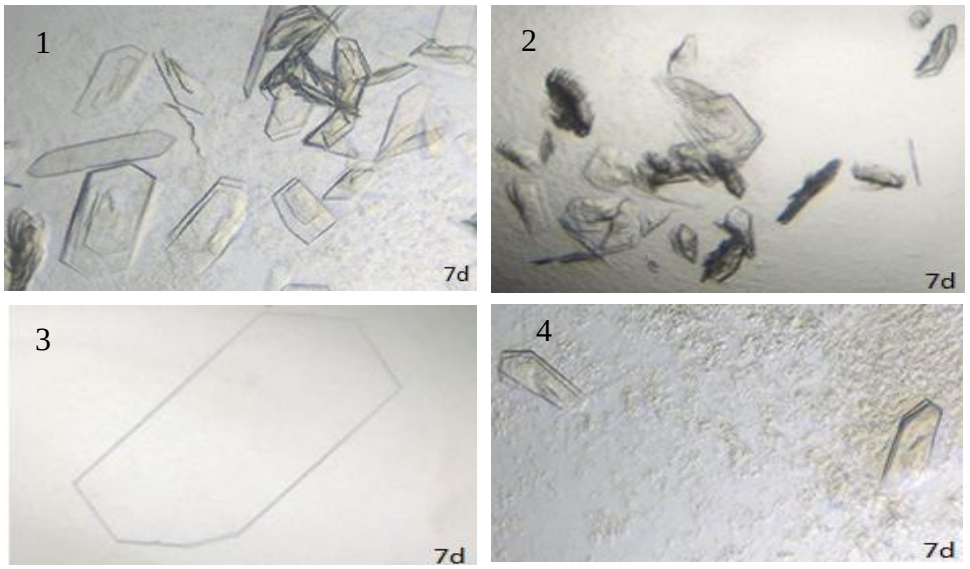


Fig IV.24 Plate-shaped crystals of HaBPS and HaBPST135L, which developed after using Poly (acrylic acid sodium salt) 2100. The numbers refer to the corresponding conditions in Table 4. The crysrals continued to grow until 7 d.

4.3. *Hypericum sampsonii* BPS

Unexpectedly, HsBPS did not show any crystal formation upon using the same screens as for HaBPS. Therefore, higher protein concentrations were used (20 mg/ml and 24 mg/ml) and many cubic shape crystals developed (Figure IV.25), when using the screens in the following Table 5.

Table 5. Buffer screen for HsBPS

Crystallization kits	Composition		Fig IV.25 panel No.
HsBPS RI screen D10 No 46	MES (0.1 M) pH 5.9 Lithium acetate 0.211 M Polyethylene glycol 3350 (15% w/v) Water	14 µl 7.389 µl 21 µl 27.611 µl	1
HsBPS R3 screen A10 No 10	MES (0.1 M) pH 5.633 Lithium acetate 0.3 M Polyethylene glycol 3350 (16.7% w/v) Water	12 µl 9 µl 20 µl 19 µl	2
HsBPS R3 screen C2 No 26	MES (0.1 M) pH 5.9 Ammonium acetate 0.367 M Glycerol (6.67% w/v) Polyethylene glycol 4000 (12.2% w/v) Water	12 µl 11 µl 4.651µl 14.667 µl 17.682 µl	3
HsBPS R3 screen C2 No 26	MES (0.1 M) pH 6.033 Lithium chloride 0.233 M Polyethylene glycol 6000 (18.3% w/v) Water	12 µl 7 µl 22 µl 19 µl	4

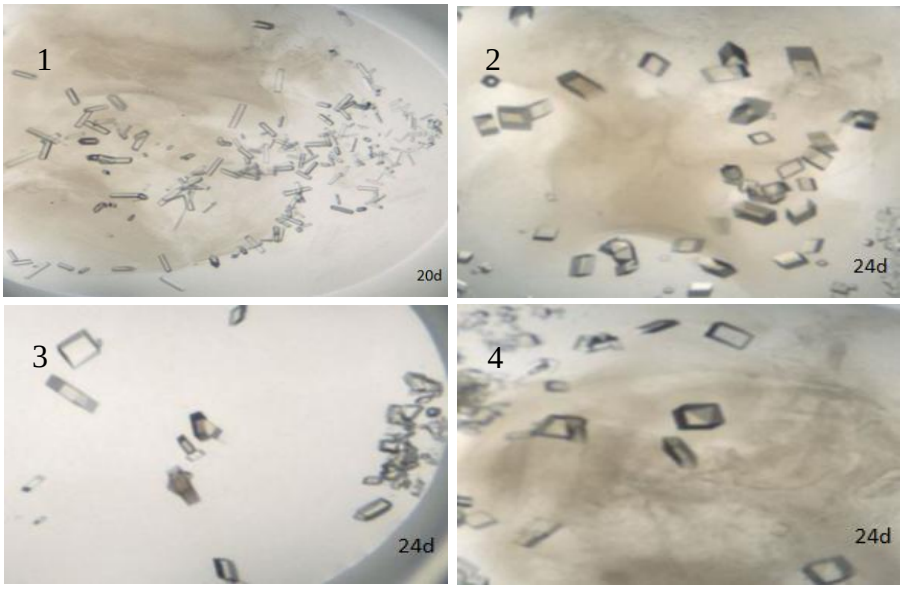


Fig IV.25 Development of cubic shape crystals of HsBPS. The numbers refer to the corresponding conditions in Table 5. The crysrals continued to grow until 24 d.

5. Enzyme substrate interactions

5.1. Isothermal titration calorimetry (ITC)

Studying the binding between the enzymes and their substrates yields information about the conditions needed for efficient interaction. This was especially true after doing the optimization efforts, which led to the use of new buffers and compoments. Therfore, it was essential to carry out ITC (III 4.5.), the results of which indicate if the tested substrates can be used for soaking and co-crystallization with the enzymes. We tested HaBPS and the mutants HaBPSC167A and HaBPSC167S, especially because of the oxidation of the catalytic Cys, which was observed in many type III PKS structures. This oxidation hinders the binding of substrates by HaBPS. The conditions tested are listed in the following Table 6, with 25 injections and 11 µl protein being used.

Table 6. Conditions used for ITC

Enzyme		Substrate		Condition
HaBPS	20 µM 40 µM 50 µM	Benzoyl-CoA	200 µM 400 µM 500 µM	50 mM Tris-HCl pH 8.5 100 mM NaCl 10% glycerol
HaBPSC167A HaBPSC167S	20 µM 20 µM		200 µM 200 µM	

Although the curves resulting from the conditions used did not give the desired information about the interaction between HaBPS and its preferred substrate benzoyl-CoA (Figure IV.26), it was not possible to test different conditions and concentrations or other substrates.

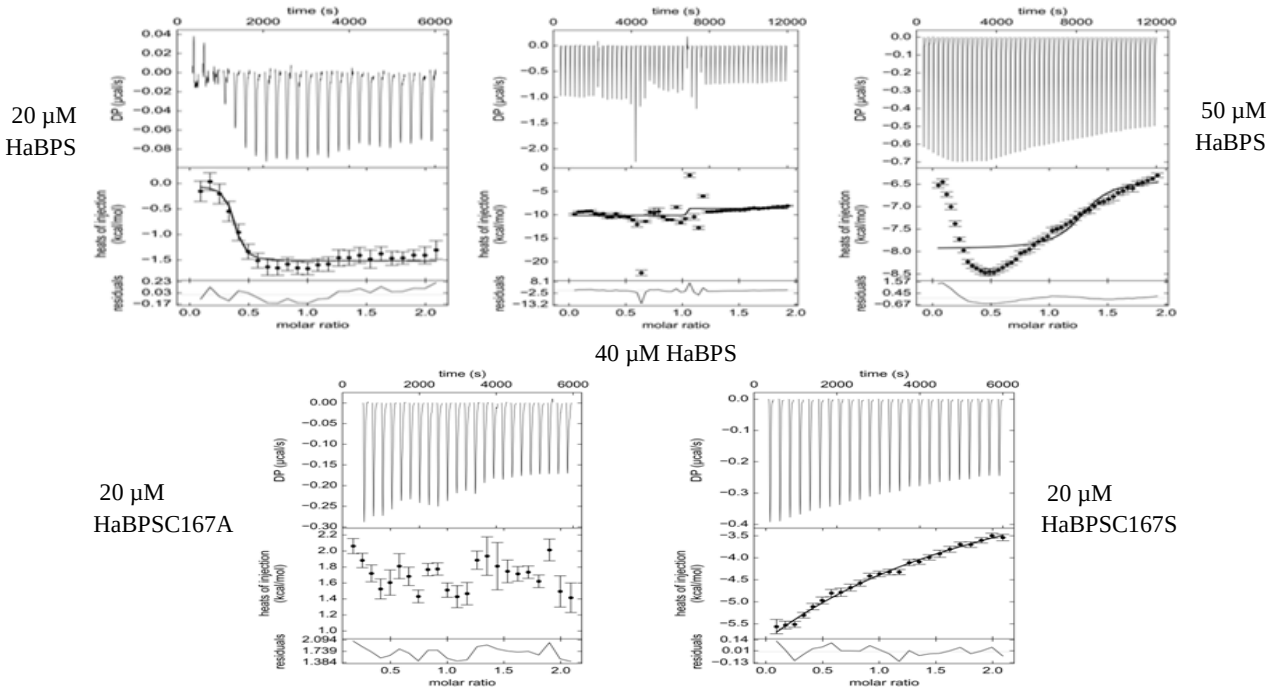


Fig IV.26 ITC of HaBPS and its active site mutants. Titration of the enzymes was performed using benzoyl-CoA as a ligand. No sigmoidal curve was observed and therfore no clear interaction was detected.

5.2. Thermal shifts assay

Since the ITC experiment did not work as expected, the implementation of an easier and faster thermal shift assay appeared rather feasible. This assay is quite similar to the thermofluor assay and, instead of using the additives, the substrates benzoyl-CoA and malonyl-CoA are added to the enzymes in their best buffer solutions (50 mM Tris-HCl pH 8.5, 100 mM NaCl and 10% glycerol). The thermal signals recorded indicate a stable interaction between HaBPS or its mutant HaBPSC167S and the substrates. This mutant was tested because its structure is closest to HaBPS; Ser is polar and similar to Cys and not hydrophobic as Ala. Thermal shifts were recorded for HaBPS and HaBPSC167S with benzoyl and malonyl-CoA and allowed for comparing their affinities to the substrates. The results indicated that substrate binding by the wild type enzyme (WT) was stronger and more stable than in its mutant. They also indicated that benzoyl-CoA is the best substrate for binding, as indicated by the higher thermal signals for HaBPS (50 °C) and HaBPSC167S (46 °C) (Figure IV.27).

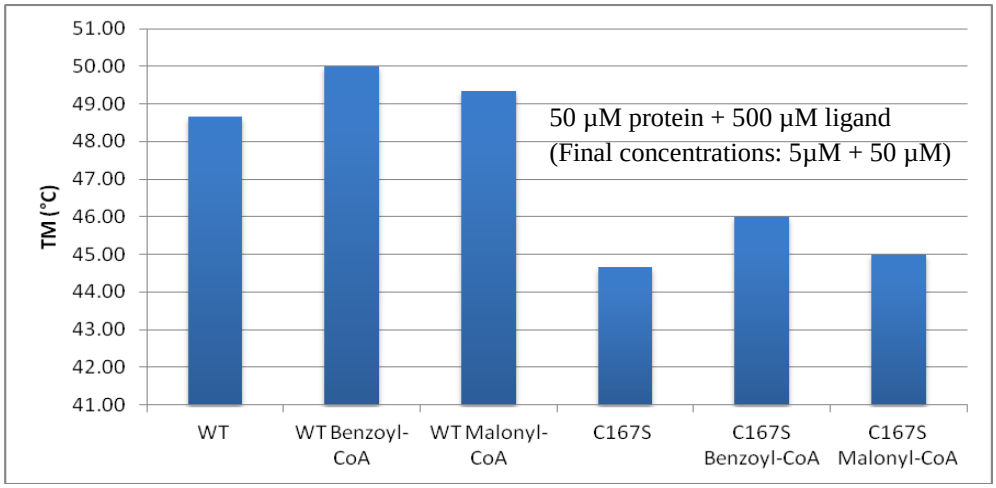


Fig IV.27 Thermal shift assay with HaBPS and HaBPSC167S. The wild-type enzyme showed higher thermal shift records with both benzoyl and malonyl-CoA in comparison to its mutant.

6. Soaking and co-crystallization

The crystals obtained for HaBPS and HsBPS as well as MdBIS3 were successfully soaked with 10 mM CoA for 5 min. Further soaking and co-crystalization were done with different substrates and products, using the wild-type enzymes HsBPS and HaBPS as well as the mutants HaBPST135L, HaBPSC167A and MdBIS3C159A. As in the mutants of HaBPS and MdBIS3, the instability of the Cys in the catalytic triad and its oxidation also appeared in some of the structure datasets of the wild-type enzymes. As an exception, only HsBPS showed unoxidized active site Cys in all of the datasets that were obtained. Therefore, this enzyme was preferably tested for soaking and co-crystallization over HaBPS and MdBIS3. The following Table 7 presents a survey of the conditions, enzymes and mutants used. Although good crystals were obtained after co-crystallization or used for soaking experiments (Figure IV.28), there was unfortunately no definite structure datasets that showed the real binding between enzyme and substrate used.

Table 7. Conditions used for substrate soaking and co-crystallization

Enzyme	Crystallization conditions		Substrate	Substrate concentration		Fig IV.28 panel NO.
				Soaking	Co-Cryst	
HsBPS 13.7 mg/ml	HsBPS R3 screen D4 No 40		Benzoyl-CoA	20 mM-16 h		1
	MES (0.1 M) pH 5.767	12 µl				
	Lithium acetate 0.333 M	10 µl				
	2-Methyl-2,4-pentanediol (4.67%w/v)	2.8 µl				
	PEG monomethyl ether 2000 (15% w/v)	18 µl				
	Water	17.2 µl				
HaBPS, HaBPS-T135L 28 mg/ml	BPS Polyacrylate screen B2 No 14 (T135L)		Benzoyl-CoA	5 mM-5 min 10 mM-3 d		2
	Tris-HCl (0.1 M) pH 7.222	6 µl				
	Poly(acrylic acid sodium salt) 2100- (30% w/v)	36 µl				
	Water	18 µl				
HaBPS-C167A 28 mg/ml	BPS Polyacrylate screen D10 No 46 (T135L)		Benzoyl-CoA	10 mM-5min 10 mM-3 d 10 mM-2 w 20 mM-2 d	10 mM	3
	di-Sodium malonate (0.161 M)	4.833 µl				
	Poly(acrylic acid sodium salt) 2100- (26.7% w/v)	32 µl				
	Water	23.167 µl				
			Malonyl-CoA	10 mM-3 d 10 mM-2 w	10 mM	
			2,4,6 tri-hydroxy-benzo-phenone	10 mM-3 d 10 mM-2 w	10 mM 20 mM	
MdBIS3-C159A 11.8 mg/ml	MdBIS3-RI screen G6 No 78 C159A		Benzoyl-CoA	20 mM-16 h 20 mM-48 h		4
	TRIS-BASE (0.1 M) pH 7.889	7 µl				
	Polyethylene glycol 4000 (21.1% w/v)	29.556 µl				
	2-Methyl-2, 4-pentanediol (8.33%)	5.833 µl				
	Lithium acetate (0.106 M)	3.694 µl				
	Water	23.917 µl				

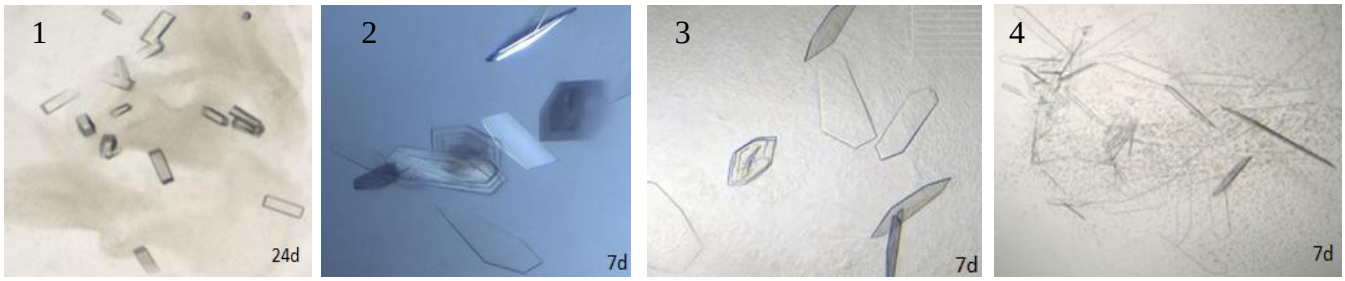


Fig IV.28 Enzyme crystals either used for substrate soaking or produced by co-crystallization. The numbers refer to the conditions used in Table 7.

7. Structure resolution

7.1. The structure of *Hypericum androsemum* BPS

Big plate-shape crystals developed for HaBPS and produced an interpretable X-ray diffraction pattern, which resulted in solving the structure of HaBPS (Figure IV.29).

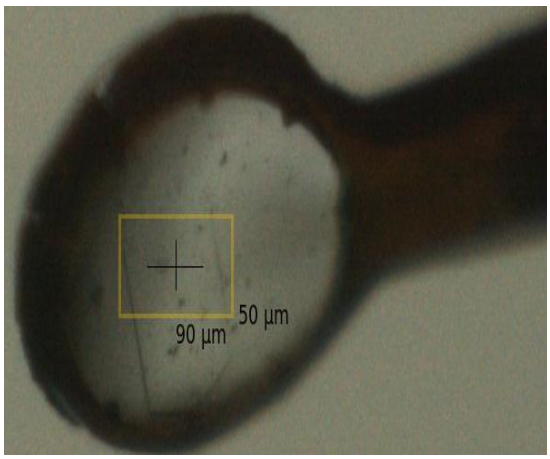


Fig IV.29 HaBPS crystal used for synchrotron measurement. The yellow-marked region represents the area that was subjected to X-ray radiation.

The crystal structure of HaBPS has the dimer shape as all type III PKSs and covers 379/409 residues, the first 30 residues of which were missing. That residues are missing from protein structures is not uncommon and happens when regions of protein molecules are disordered in the protein crystal and therefore can not generate a good diffraction pattern (Figure IV.30).

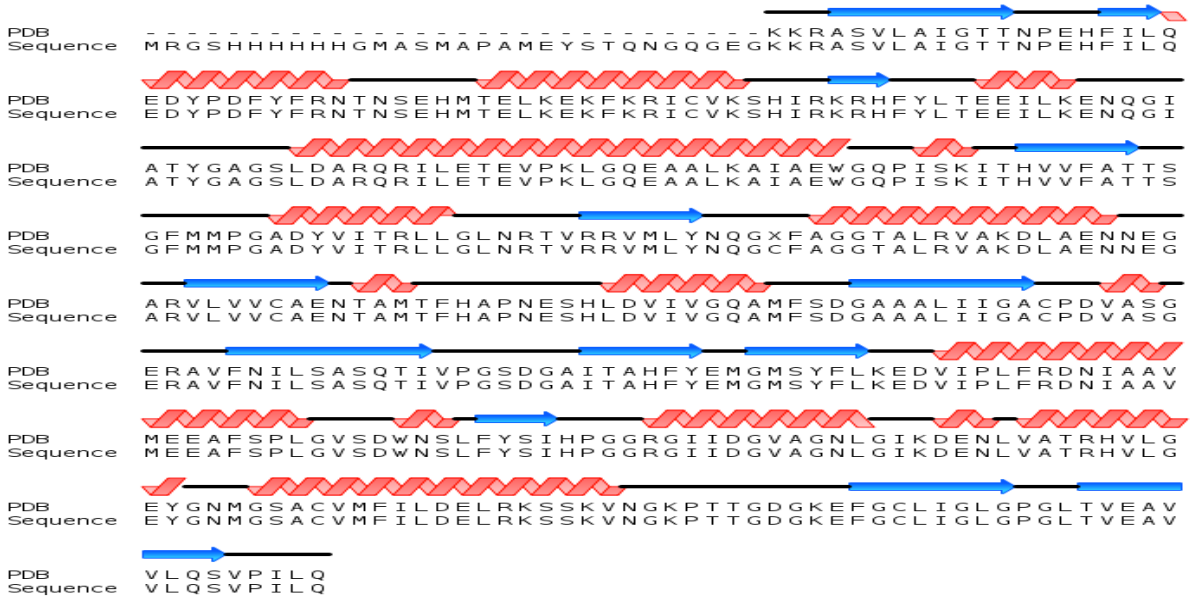


Fig IV. 30 Secondary structure of HaBPS. It lacks the 30 N-terminal residues, the first 14 residues of which do not belong to the protein but to the His_6-tag region.

The crystal structure of HaBPS consists of two main domains, the upper domain and the lower domain. The lower domain is more important, as most of the active site pocket is located in this area. The active site entrance is also located at the bottom, which presents the main door for the entry of the substrates toward the active sites cavity (Figure IV.31).

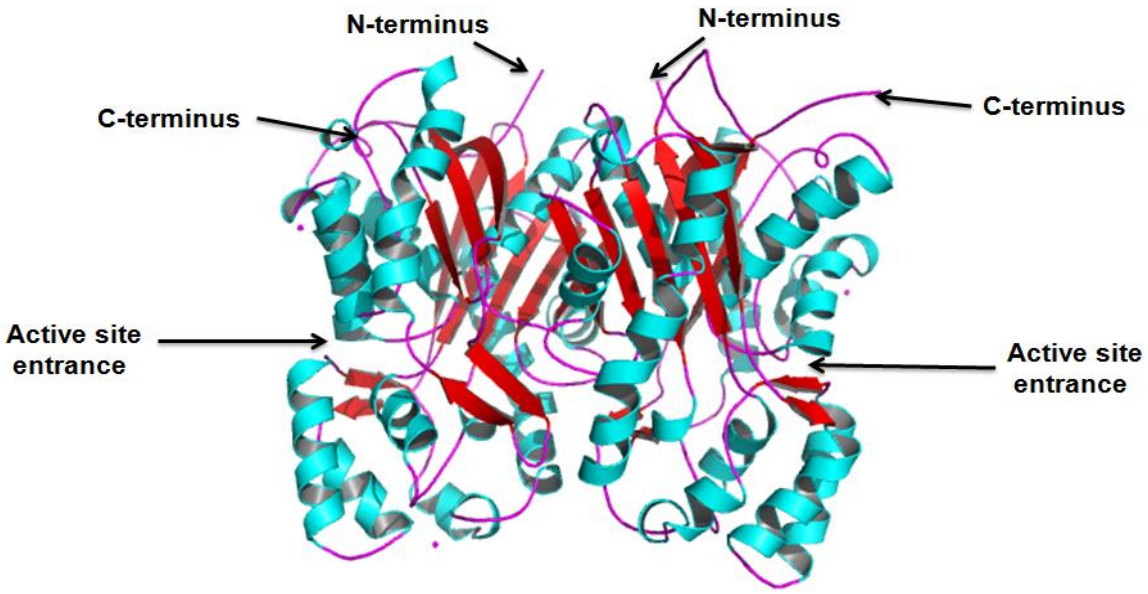


Fig IV.31 The crystal structure of HaBPS with the αβαβα domain typical of type III PKSs.

The structure of HaBPS was solved at 2.1 Å by molecular replacement (MR) using chalcone synthase from *Freesia hybrid* (FhCHS), which has recently been solved and shares 62% sequence identity with HaBPS at the amino acids level (Figure IV.32). With MsCHS, HaBPS shares 57.1% identity.

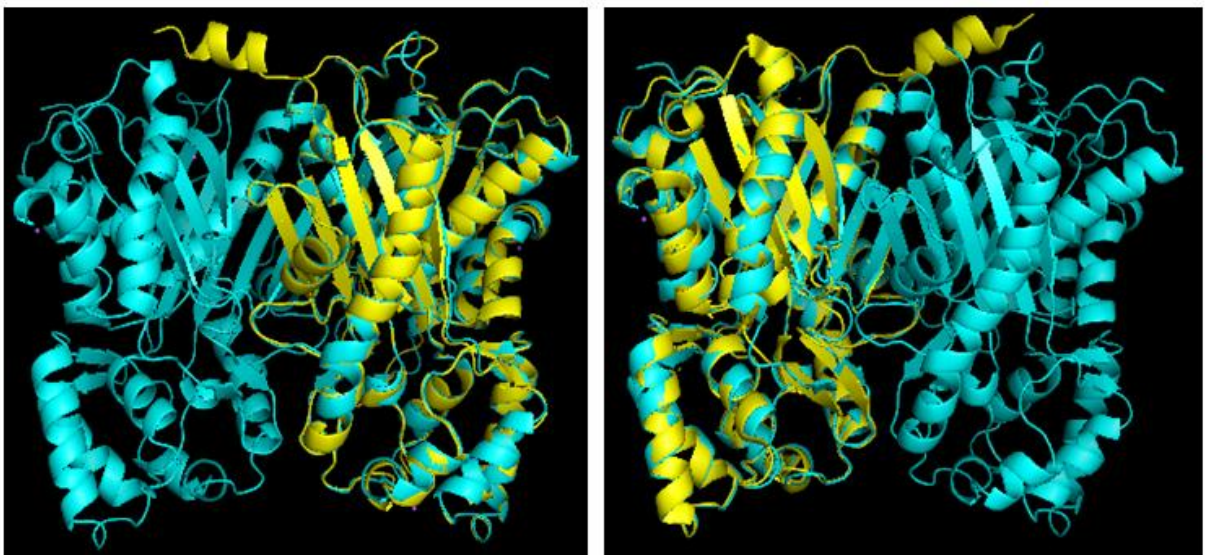


Fig IV.32 Superimposed structures of HaBPS (yellow) and FhCHS (blue), which share 62% identity.

The structure of HaBPS in complex with CoA was solved at 1.5 Å. It shows the orientation of the CoA molecule in interaction with the residues of the CoA-binding tunnel (Figure IV.33).

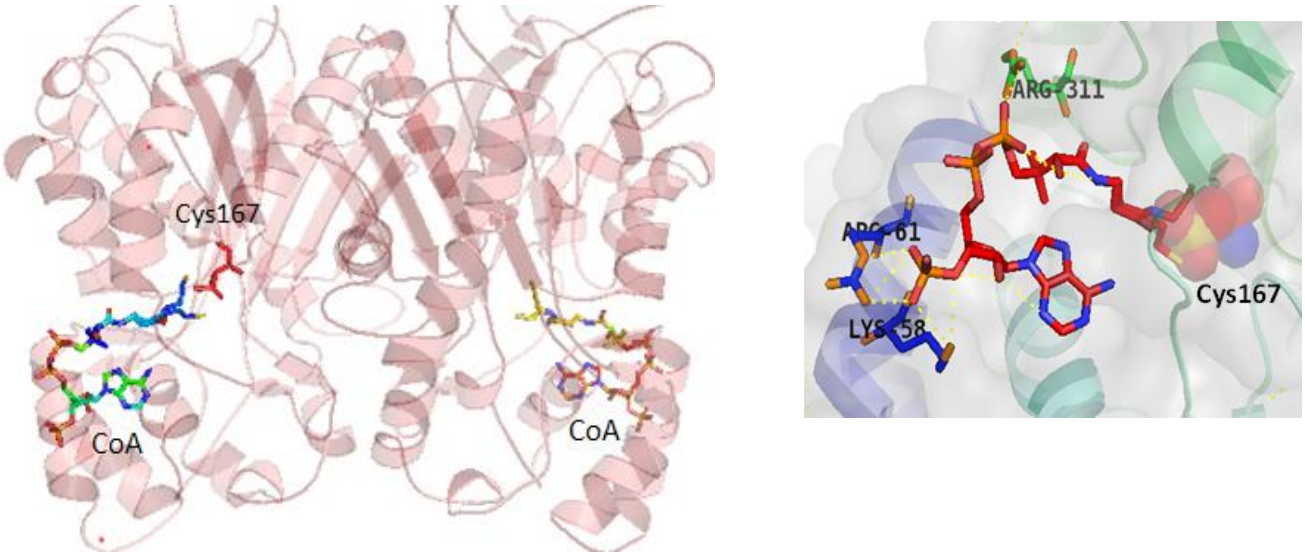


Fig IV.33 Crystal structure of HaBPS in complex with CoA. The CoA binding tunnel (right) contains the essential residues at the active site entrance (Arg 61, Arg 311 and Lys58), which interact with CoA by hydrogen bonds (yellow dotted lines). The essential catalytic residue Cys167 is shown as spheres.

7.2. The structure of *Hypericum androsaemum* BPST135L

The structure of the T135L mutant was solved at 1.4 Å, covering 378/409 residues, the first 31 residues of which were missing (Figure IV.34).

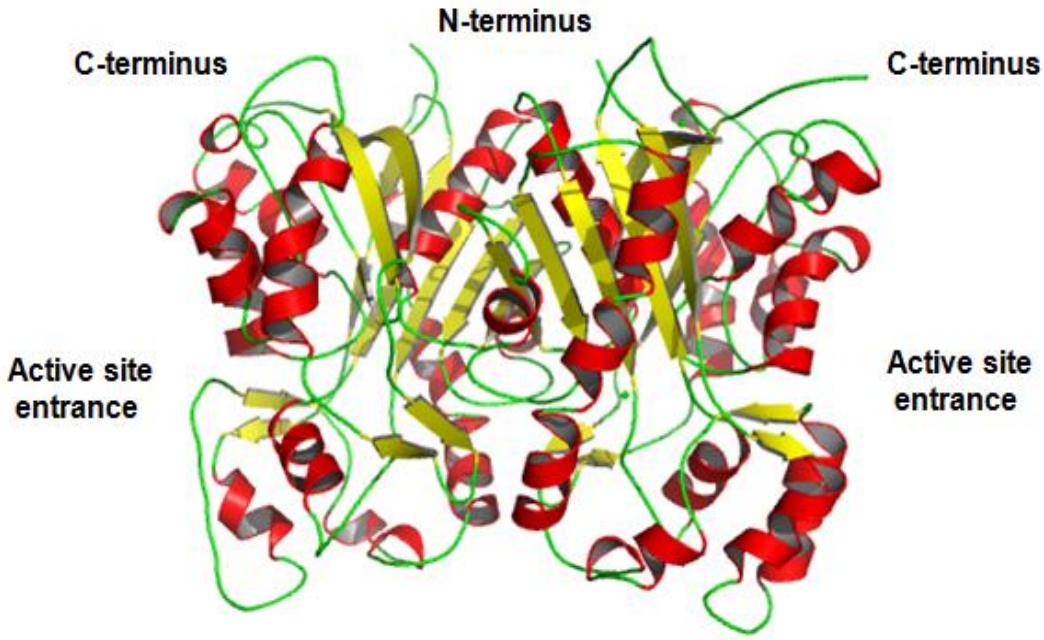


Fig IV.34 The crystal structure of HaBPST135L with the αβαβα domain typical of type III PKSs.

The structure of HaBPST135L shows the proper geometry of the residues surrounding the mutation site. The orientation of Leu135 proves the formation of a new pocket, as explained previously by homology modelling of the mutant. It explains also why this mutant is inactive with 3-hydroxybenzoyl CoA, as this substrate mimiks the function of the absent Thr135 and forms hydrogen bonds with Gly166, resulting in trapping the substrate inside the new pocket (Figure IV.35) (Klundt et al., 2009).

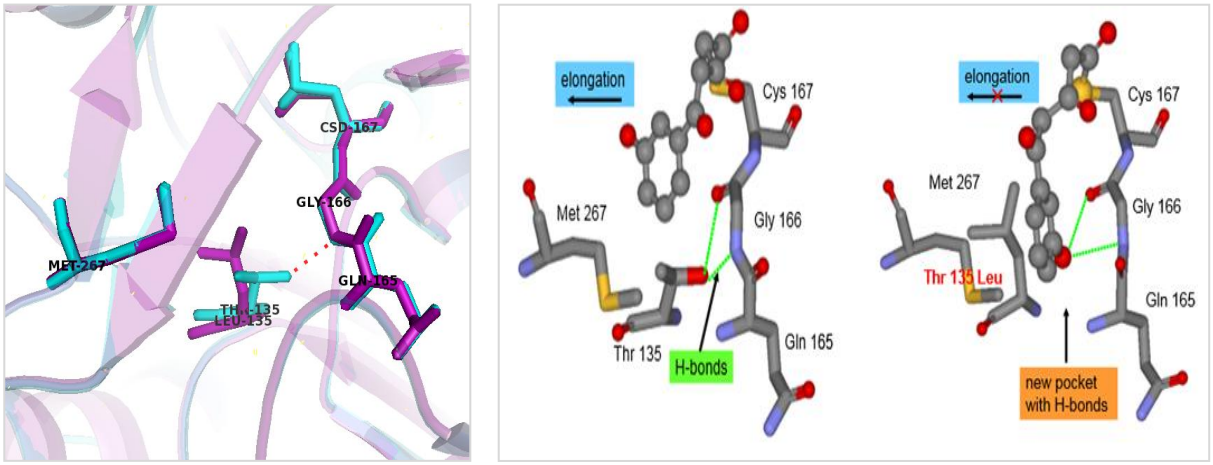


Fig IV.35 Different orientations of T135 in HaBPS (blue) and L135 in HaBPST135L (purple). On the right, trapping of the 3-hydroxybenzoyl primed intermediate in the newly formed pocket blocks the further elongation reaction (Klundt et al., 2009).

The size of the active site pocket has a large impact on the number of condensations with malonyl-CoA and the length of the polyketide chain. This holds true for the active site pockets of wild-type BPS, which condenses three molecules of malonyl-CoA, and its mutant T135L, which condenses only two molecules of malonyl-CoA in the new pocket (Figure IV.36).

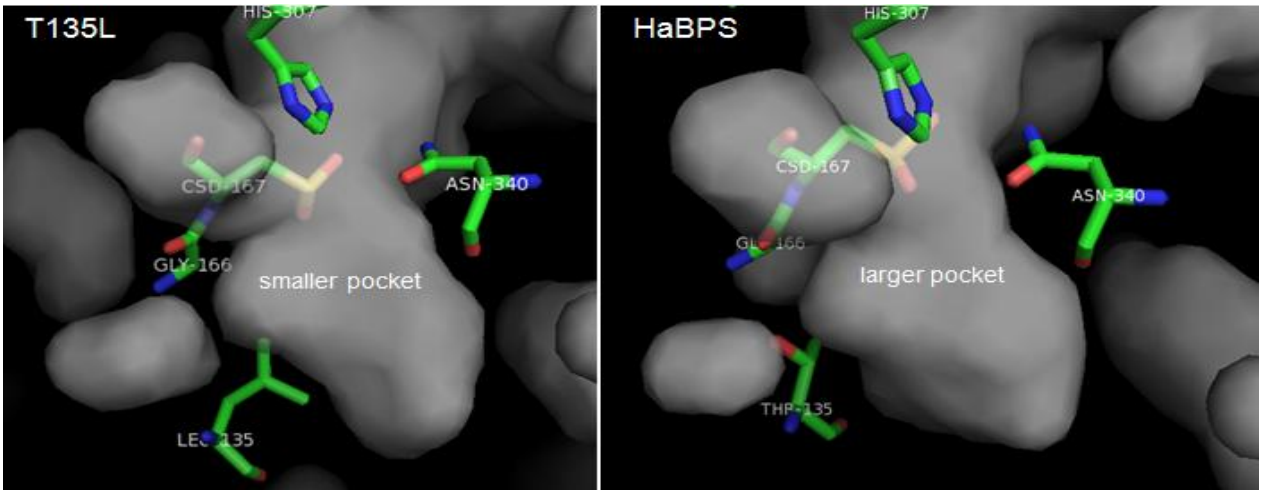


Fig IV.36 The size of the pockets surrounding Leu135 in HaBPST135L and Thr135 in HaBPS. The active site pocket in HaBPST135L is smaller than in HaBPS, explaining the production of a triketide chain by the mutant rather than a tetraketide chain by the wild-type enzyme.

7.3. The structutre of *Hypericum androsaemum* C167A

Good crystals of HaC167A were used for soaking benzoyl- and malonyl-CoAs and the product 2,4,6-trihydroxybenzophenone. Good diffraction patterns were obtained and the structure of HaC167A was solved at 2.6 Å. However, when the active site pocket was analyzed there was no complex formation in spite of the presence of Ala167, which widened the CoA-binding channel (Figure IV.37).

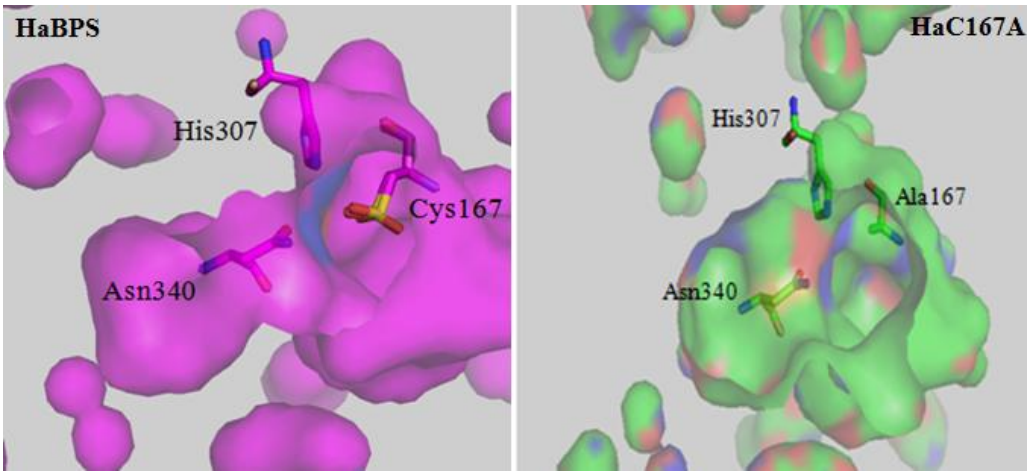


Fig IV.37 Comparison of the CoA-binding tunnels in HaBPS and HaBPSC167A. On the right, the view through the CoA-binding tunnel in C167A presents Ala instead of oxidized Cys at position 167. The prescence of Ala made the channel wider in comparison to HaBPS on the left.

7.4. The structure of *Hypericum sampsonii* BPS

The crystals of HsBPS had a nice cubic shape, although smaller than those of HaBPS. After synchrotron measurement, the structure of HsBPS was solved at 1.7 Å by molecular replacement using HaBPS, in which the first 30 residues were missing (Figure IV.38).

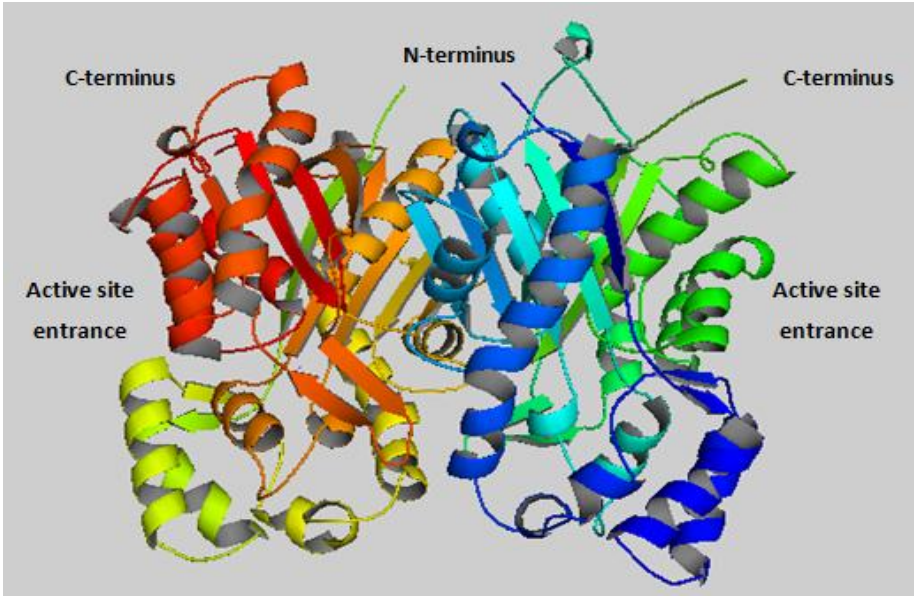


Fig IV.38 The crystal structure of HsBPS with the αβαβα domain typical of type III PKSs.

There are only six amino acids in the structure of HsBPS, which differ from HaBPS. They are lying on the surface of the enzyme (Figure IV.39).

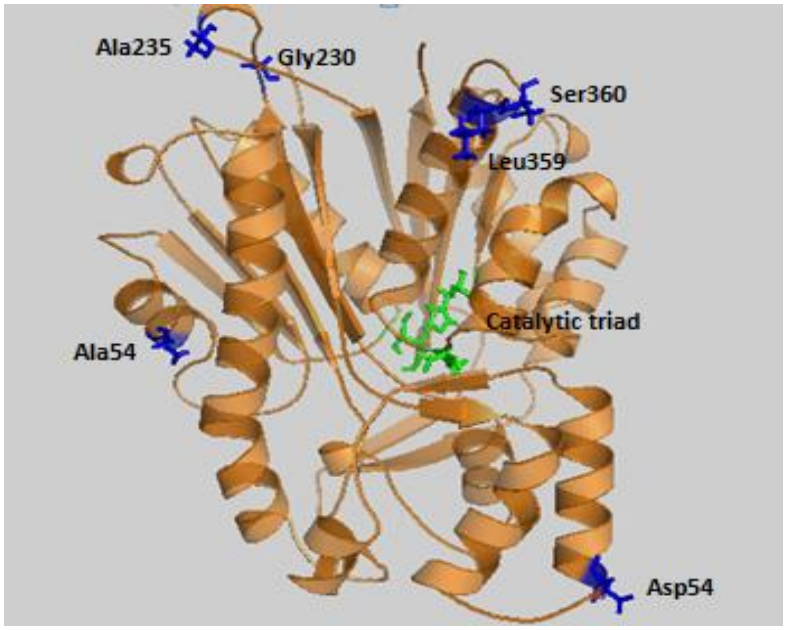


Fig IV.39 The crystal structure of HsBPS. The catalytic triad (green) and the six amino acids that differ from HaBPS (blue) are highlighted.

The six different amino acids did not only change the biochemical properties (Abdelaziz, 2014) but also the active site pocket, which differed slightly in shape between HsBPS and HaBPS (Figure IV.40)

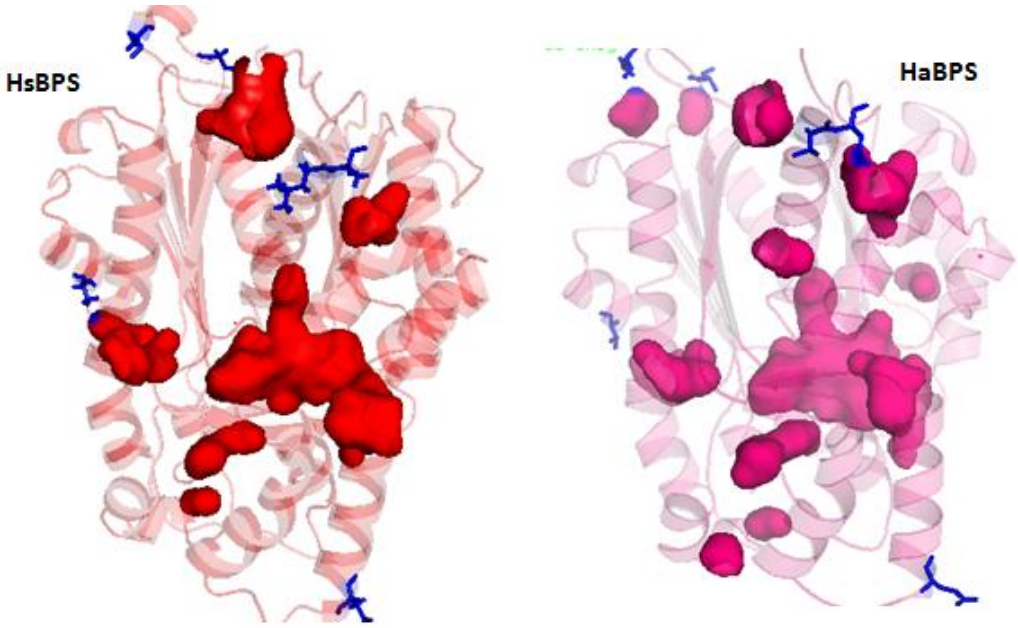


Fig IV.40 Difference in the active site pockets of HaBPS (red) and HsBPS (purple). The six different amino acids, which are located on the surface, are highlighted as sticks (blue).

7.5. The structure of *Malus domestica* BIS3

The structure of MdBIS3 was solved at 1.9 Å. Of the 380/402 residues, the first 22 residues were missing (Figure IV.41).

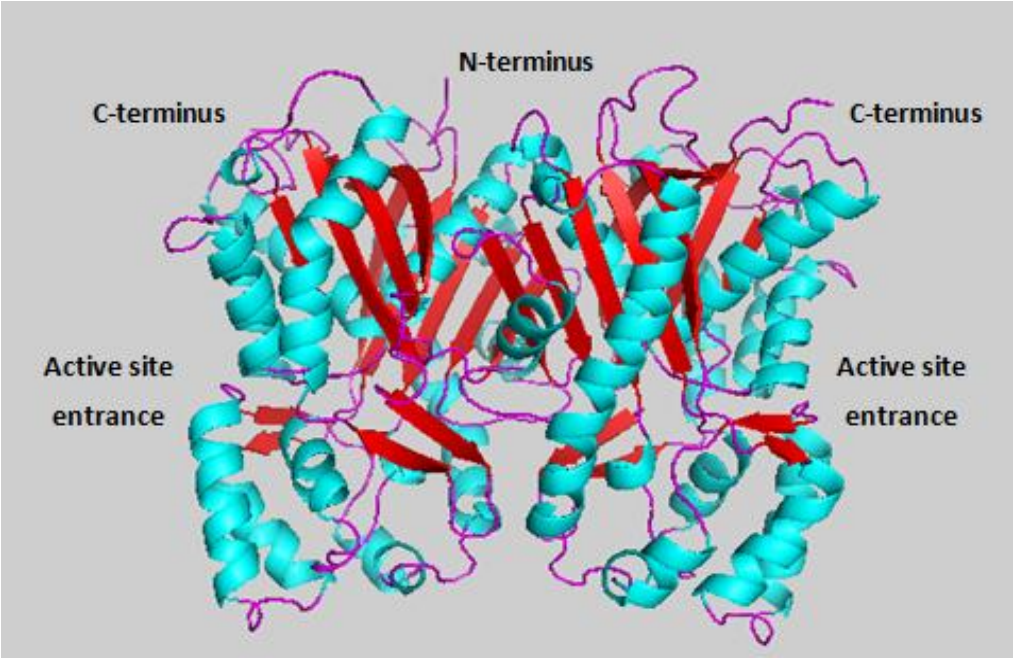


Fig IV.41 Crystal structure of MdBIS3 with the αβαβα domain typical of type III PKSs.

Crystals were soaked for 48 h in 20 mM fresh benzoyl-CoA solution and data were collected at a resolution of 2.5 Å. Co A is clearly present in two of the four monomers with occupancy of 70-80% (Figure IV.42).

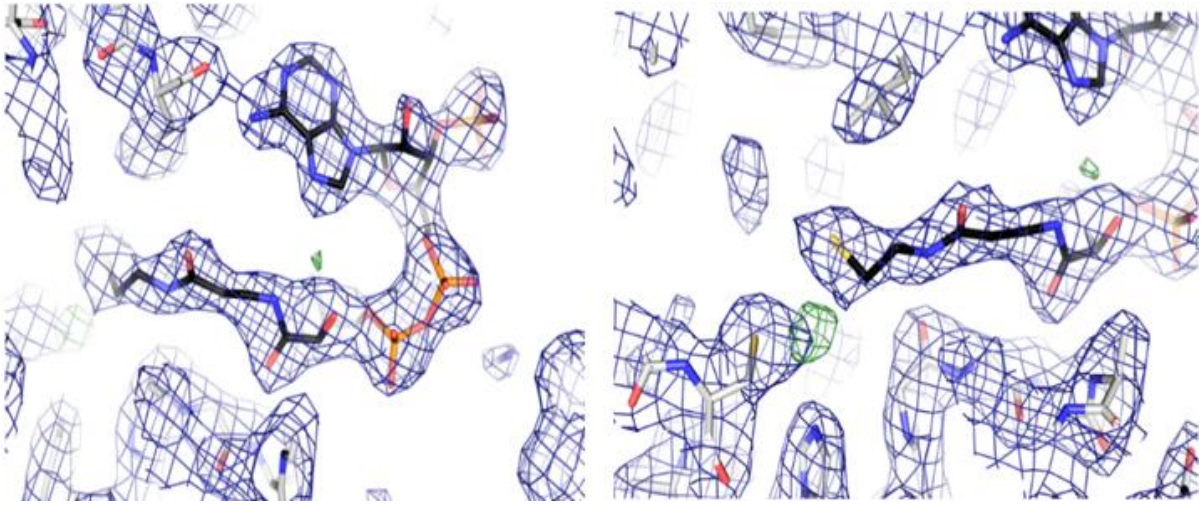
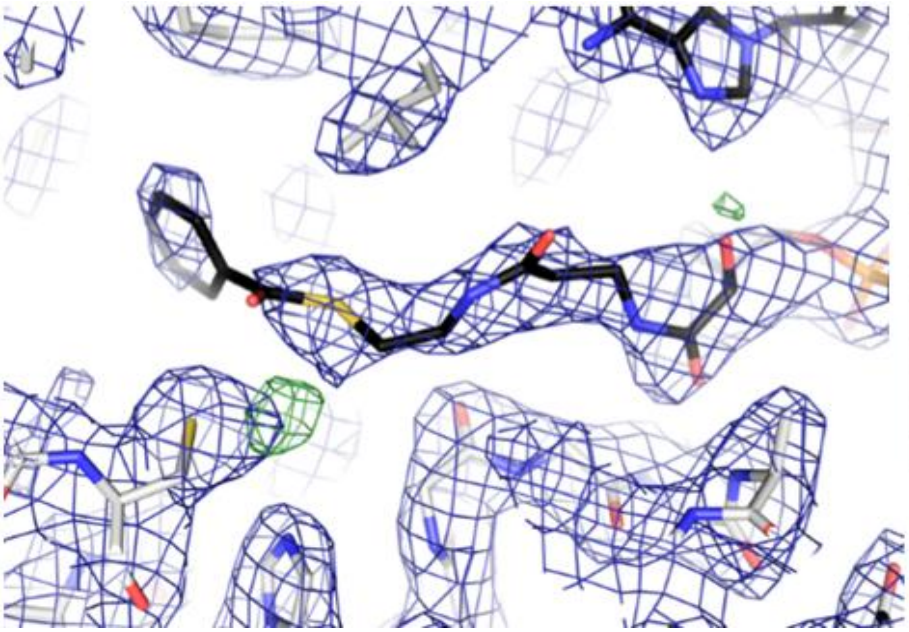


Fig IV.42 Presence of CoA molecules in two monomers of MdBIS3.

For the first time, we could see an electron density that fit perfectly to the benzene ring close to the catalytic Cys and we were able to see the whole benzoyl-CoA molecule (Figure IV.43).



CoA

	Starting	Final
R-work	0.1865	0.1843
R-free	0.2192	0.2160
Bonds	0.012	0.004
Angles	1.011	0.673

Benzoyl-CoA

	Starting	Final
R-work	0.1826	0.1813
R-free	0.2181	0.2175
Bonds	0.010	0.004
Angles	1.001	0.681

Fig IV.43 Positive electron density of benzoyl-CoA (blue mesh) showing high Rf values.

The positions of the neighboring residues showed perfect orientation with the location of benzoyl-CoA in the active site pocket (Figure IV.44).

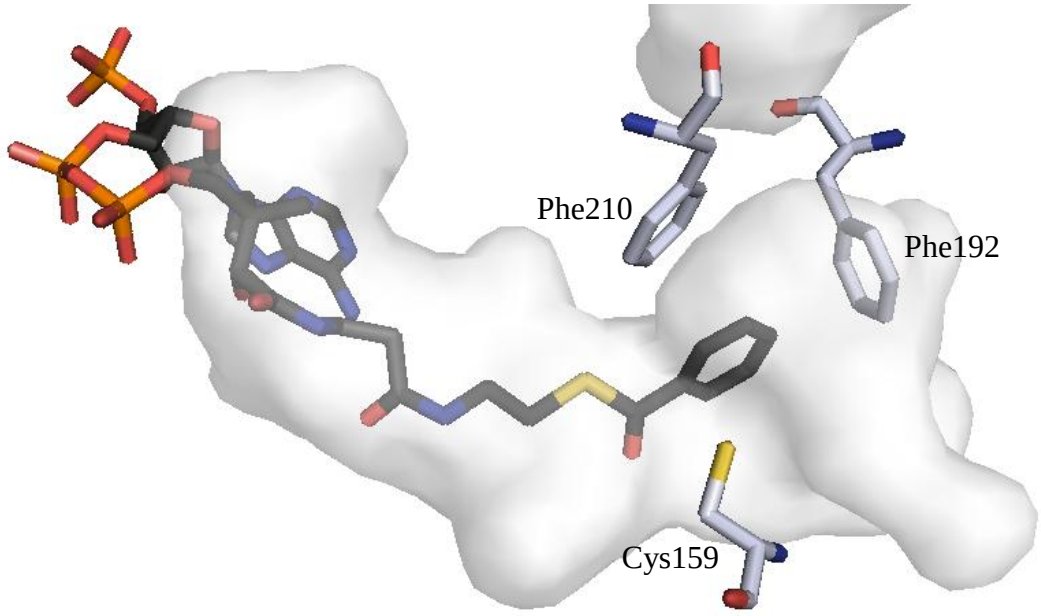


Fig IV.44 Location of benzoyl-CoA in MdBIS3. The orientation of the neighboring residues in the active site cavity (Phe210 and Tyr64) is consistent with the position of the benzene ring in the benzoyl-CoA molecule.

The active site pocket of MdBIS3 strongly differs from that of HaBPS. The shape and the size of each pocket indicate why the enzymes catalyse different reaction mechanisms (Figure IV.45a, b).

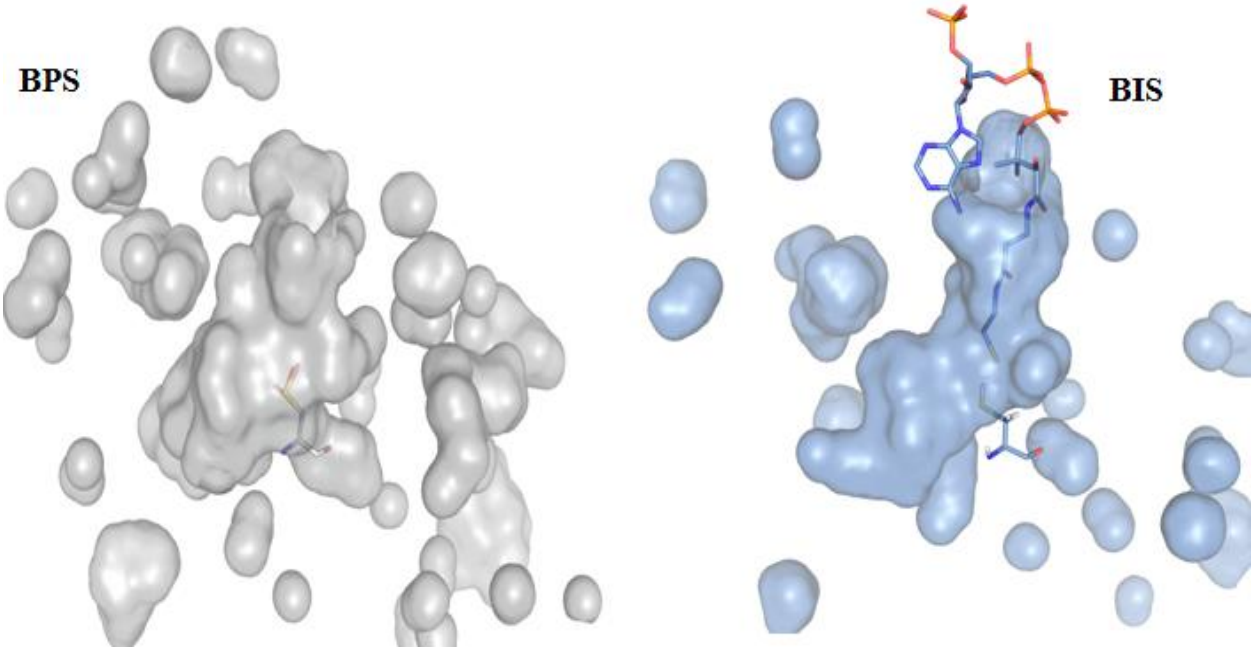


Fig IV.45a Comparison between the active site pockets of HaBPS (gray) and MdBIS3 (blue). The Cys of the catalytic triad is shown as sticks. The CoA molecule is displayed as sticks in MdBIS3 only.

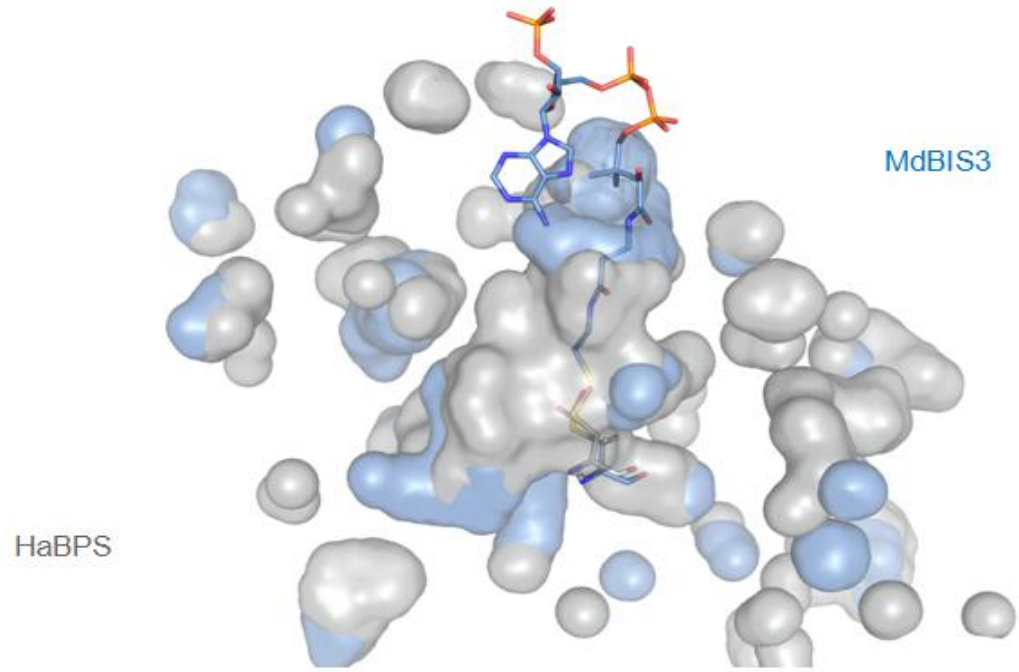


Fig IV.45b Superimposed structures of HaBPS (gray) and MdBIS3 (blue) active site cavities. The catalytic Cys and the CoA molecule are shown as sticks.

7.6. The structure of *Malus domestica* C159A

For the first time, the substrate benzoyl-CoA was seen inside the active site cavity of BIS3. Therefore, soaking trials were carried out with the active site mutant C159A, as the cysteine was still not stable and underwent again oxidation in one of the chains of the MdBIS3 structure.

The crystals of C159A were soaked with 20 mM 3,5 dihydroxybiphenyl for 48 h and the structure was solved at 1.86 Å. However, after structure data analysis the product was not found inside the active site cavity, even with the widened channel (Figure IV.46)

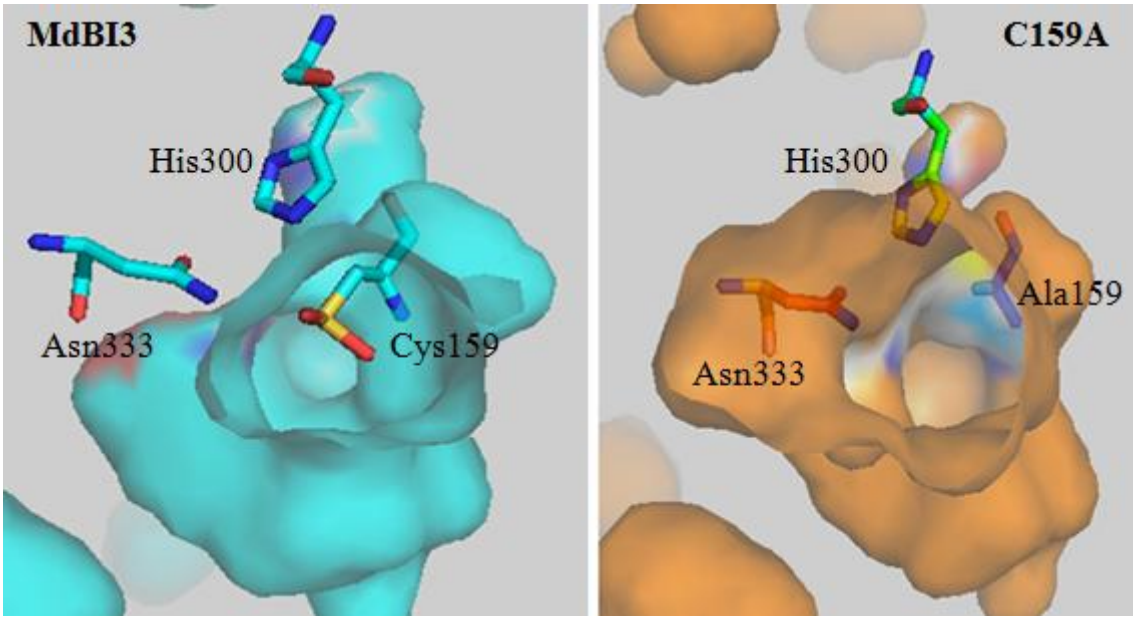


Fig IV.46 Comparison of the CoA-binding tunnels in MdBIS3 and MdBIS3C159A. On the right, the view through the CoA-binding tunnel in C159A presents Ala instead of oxidized Cys at position 159. The prescence of Ala made the channel wider in comparison to MdBIS3 on the left.

8. Transformation of BPS into BIS by mutation

Mutation of BPS toward BIS was an idea to identify which residues are responsible for the change in the cyclization mechanisms, i.e. from Claisen to aldol condensation. Preliminary understanding of the differences in the reaction mechanisms was expected from the enzyme structures in complex with substrates and products after soaking or co-crystallization. Because this was not feasible, the idea of mutation emerged.

8.1. Depending on *Sorbus aucuperia* BIS1

In the beginning, there was only the structure of HaBPS but not yet of MdBIS3. Thus, we used only the structure data of SaBIS1, which was solved previously by Raeth (2007). These were the only available data about the structure of this enzyme. We started with a 14-fold mutant depending on the differences in the side chains of the residues lining the active site cavities of HaBPS and SaBIS.

A synthetic gene for HaBPS containing the 14 mutations was ordered (Figure IV.47a, b).

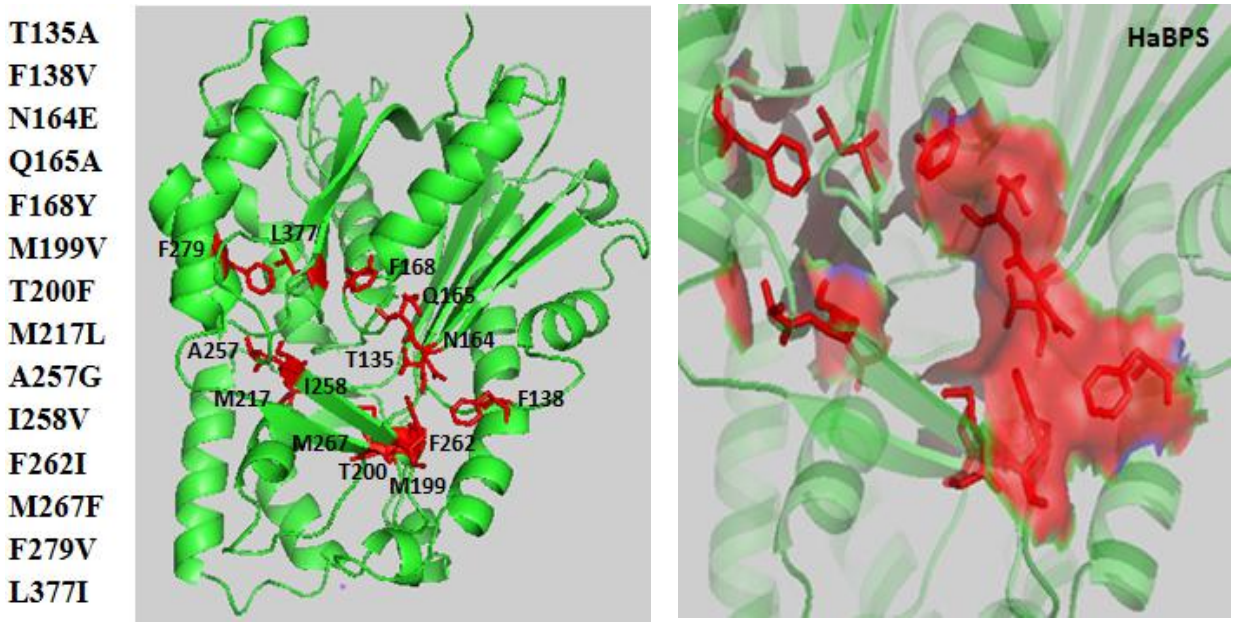


Fig IV.47a Structure of HaBPS with the 14 residues selected for mutation. The mutation sites are shown as red sticks (left). The shape of the active site pocket (red) includes the sites for mutation (right).

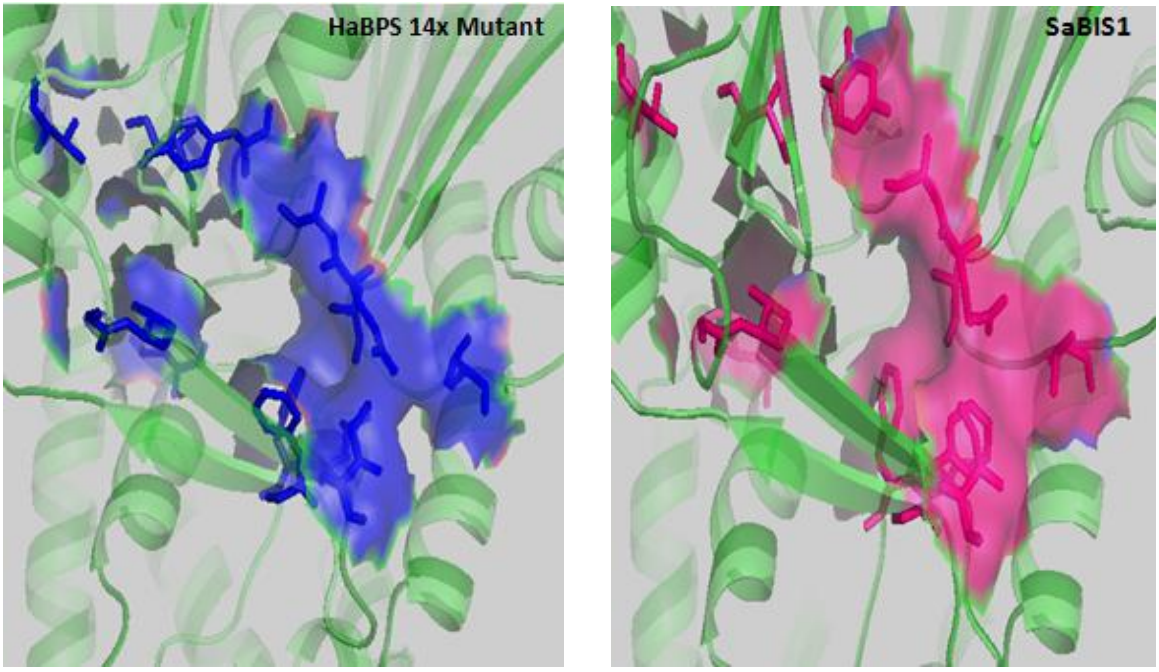


Fig IV.47b Comparison between the active site pockets in the modeled HaBPS 14-fold mutant (blue) and SaBIS1 (pink). The mutation sites are shown as sticks.

Unfortunately this 14-fold mutant showed neither BPS nor BIS activity.

8.2. Depending on *Malus domestica* BIS3

After comparison between the residues of the active site cavities of HaBPS and MdBIS3, a new 13-fold mutant was produced starting from the previous 14-fold mutant. Four residues were mutated back to their original amino acids and three new mutations were added. So, in total seven mutation sites were changed (Figure IV.48a, b).

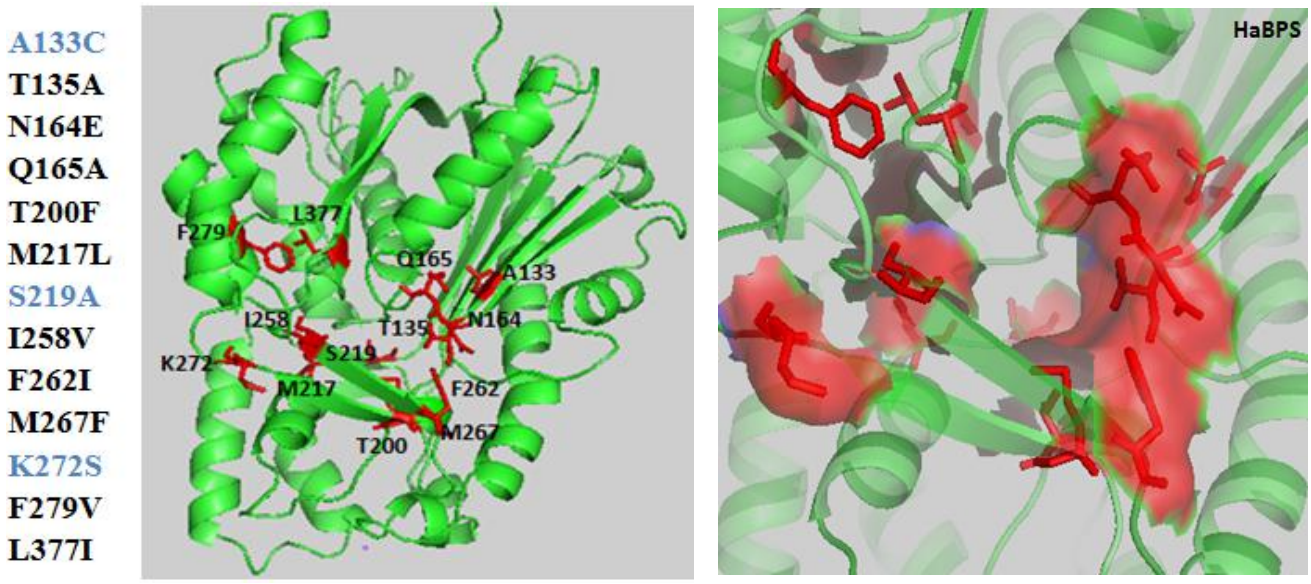


Fig IV.48a Structure of HaBPS with the 13 residues selected for mutation. The mutation sites are shown as red sticks (left). The shape of the active site pocket (red) includes the sites for mutation (right).

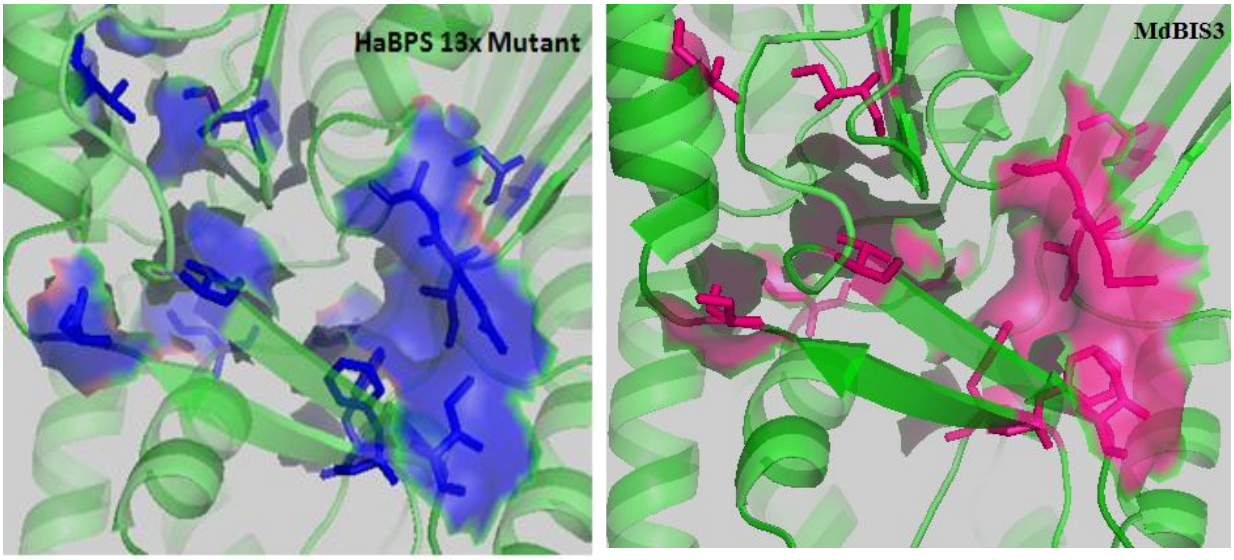


Fig IV.48b Comparison between the active site pockets in the modeled HaBPS 13-fold mutant (blue) and MdBIS3 (pink). The mutation sites are shown as sticks

Unfortunately, this 13-fold mutant also lacked catalytic activity. Further mutations were not tackled.

V. Discussions

1. Crystallization of type III PKSs

1.1. With and without purification tag

All crystallized type III PKSs were purified with the help of a purification tag, which is either a His-tag as in MsCHS, PsSTS, GhPS, CmACS, CmQNS, OsCUS and AaOKS or a GST-tag as in RpBAS and AaPCS. The GST-tag in RpBAS and AaPCS was removed by protease digestion. The His-tag sizes and locations were different among the enzymes, for example, it is His_8-tag in MsCHS, His_6-tag in CmACS and His_4-tag in ClCURS. Its location was mostly at the N-terminus, except for AaOKS and ClCURS (Ferrer et al., 1999, Morita et al., 2007, Abe et al., 2007b, Morita et al., 2010b, Mori et al., 2013). Very recently, the structures of MdBIS3 and HaBPS (derived from a synthestic gene) have been solved and both enzymes have been digested with thrombin protease to remove a His_8-tag before the crystallization process (Stewart et al., 2017). A similar digestion of His_6-tag using thrombin was carried out earlier in our group and the crystallization process of SaBIS1 was successful. However, the collected datasets were not of good resolution to get a clear structure of the enzyme (Raeth, 2007). Our target enzymes (HaBPS, HaBPST135L, HsBPS and MdBIS3) were crystallized as His_6-tagged proteins without cleavage because digesting them using TEV protease was unsuccessful. Digestion with TEV protease appears not to be the right choice for our enzymes, which may explain why the previous crystallization work, which mainly depended on the cutting of the His_6-tag using TEV protease, did not advance to the desired results. This is clearly seen by SDS PAGE gel comparison. The digestion with thrombin is complete and resulted only in one band but TEV protease digestion was incomplete and resulted in a second band (Figure V.1) (Raeth, 2007, Lütge, 2012).

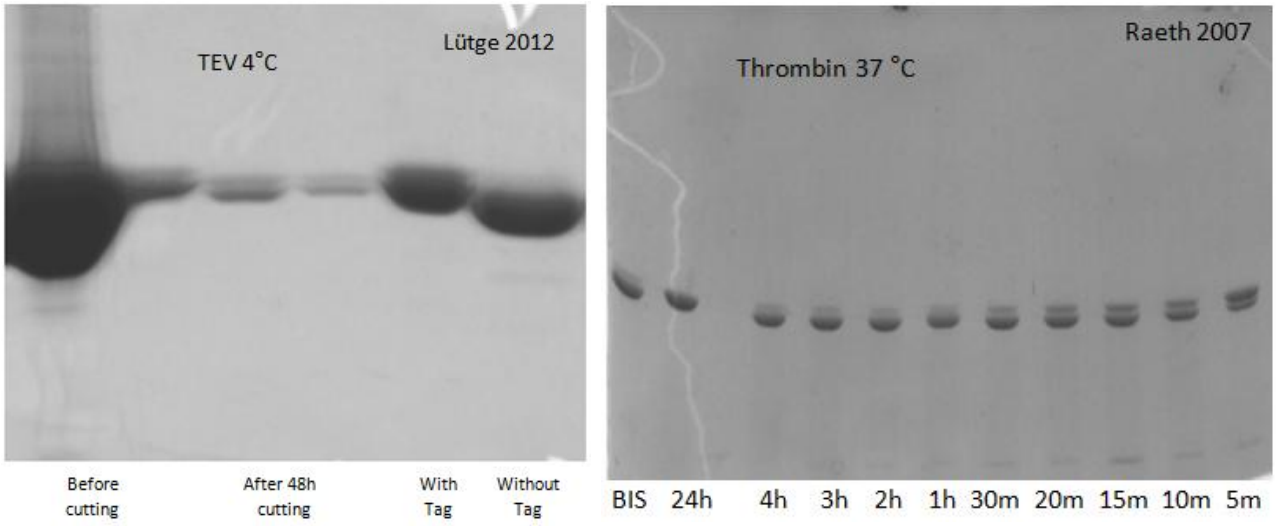


Fig V.1 The best conditions for His-tag cleavage from SaBIS1 when regarding previous crystallizations works. In 2007, Raeth optimized the cutting using thrombin at 37 °C for only 4 h. In 2012, Lütge optimized the cutting using TEV protease at 4 °C and for 48 h. The cutting using TEV protease is incomplete and shows a faint second band in the finally digested protein sample.

We followed the same cutting protocol of Raeth (2007) but used TEV protease instead of thrombin. Although the cutting was complete by SDS-PAGE analysis, the gel filtration chromatography showed the appearance of another band (Figure V.2).

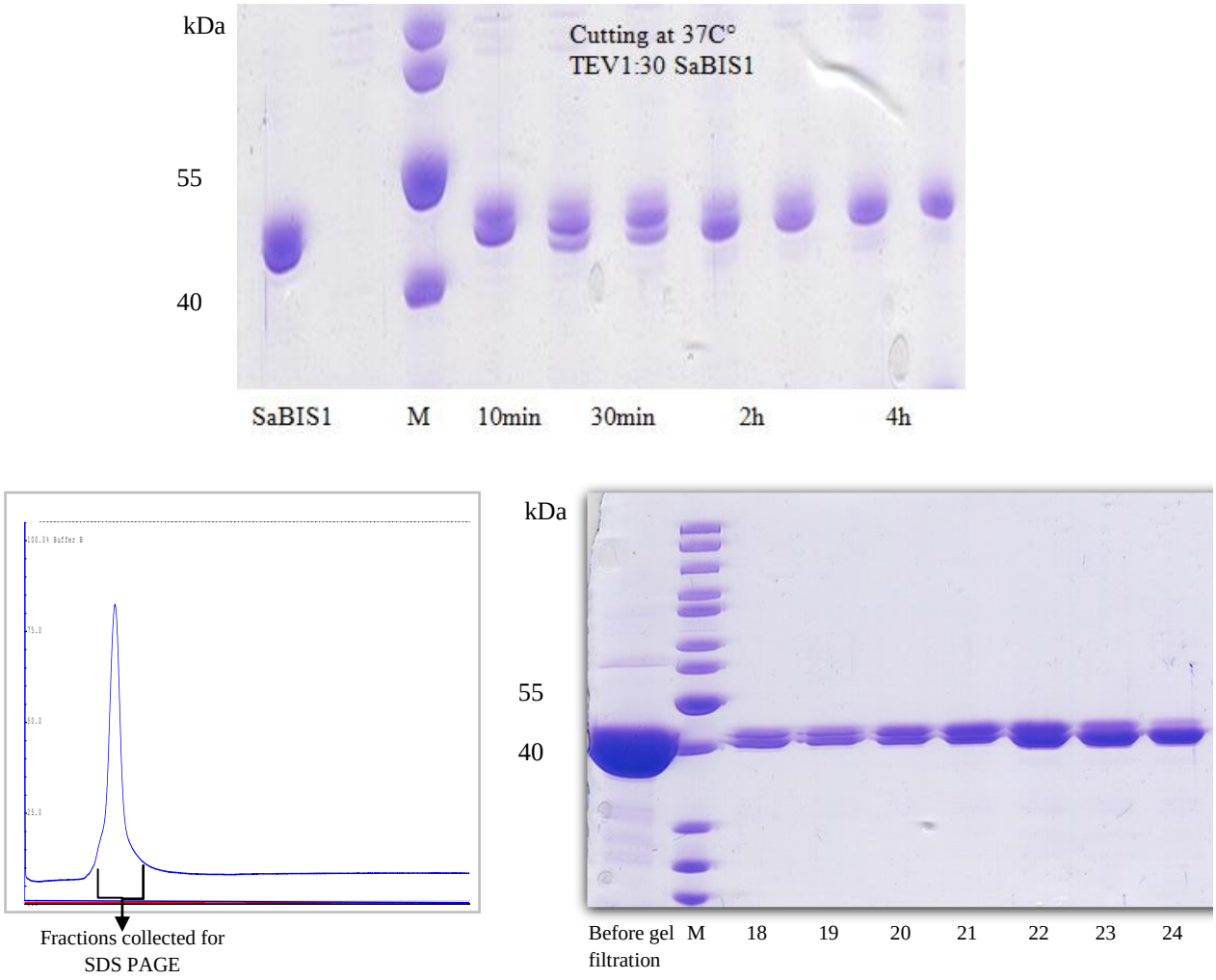


Fig V.2 The inefficient cutting of SaBIS1. The upper SDS PAGE shows the complete cutting of His_6-tagged SaBIS1 using TEV protease. The lower picture shows a sharp peak after gel filtration chromatography of the digested sample, but SDS PAGE of the final corresponding fractions (indicated by numbers) shows a mixture of two bands.

The second faint band of the non-cutted enzyme indicates that its amount is less in the final protein sample but its presence was like an impurity. Therefore, the final sample containd non-homogenous protein species. As the two bands were so close in size to each other, the final protein sample always showed one sharp peak in gel filtration chromatography. Even when the location of the N-terminal His_6-tag in SaBIS1was changed to a C-terminal His_6-tag or when a short construct of HaBPS was used (BPS-19AA), the digestion was always incomplete.

Because of the previously mentioned complications about the non-convenient cutting with TEV protease, we suggested that the N-terminal and C-terminal parts of the enzymes studied here are folded in a pattern that makes them inaccessible to TEV protease. Therefore, we proceeded using the purified but still His_6-tagged enzymes for the crystallizations processes.

1.2. The catalytic triad and oxidized cysteine

The essential three amino acids for the activity of all type III PKSs are included in the catalytic triad (Cys164, His303 and Asn336) (numbering from MsCHS2). Among these residues Cys is in the top, because the absence or the mutation of this site abolishes the enzyme activity (Lanz et al., 1991). The importance of this residue appears from its nucleophilic character, which provides an efficient binding site for the starter and the polyketide intermediates. Particular mutation of Cys164 to serine or alanaine reduces but maintains the decarboxylation of malonyl-CoA (Jez et al., 2000b). To understand the different reaction mechanisms of BPS and BIS, a clear picture about the interaction between this Cys in the enzymes and their substrates was needed. Therefore, many ligands were used for soaking into the produced crystals or for co-crystallization with the enzymes, such as benzoyl-CoA, malonyl-CoA, 3,5 dihydroxybiphenyl and 2,4,6, trihydroxybenzophenone at different concentrations and time intervals, as mentioned in IV.7. None of these trials were successful, which was correlated to the oxidized state of the catalytic cysteine, as clearly detected in the structures of HaBPS and HaBPST135L (Figure V.3).

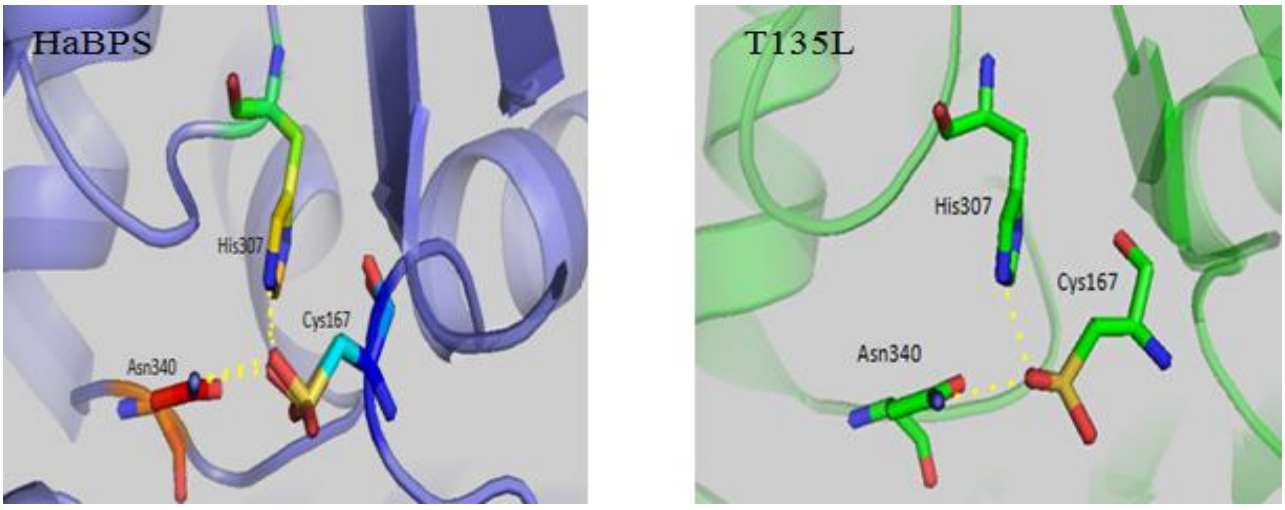


Fig V.3 The oxidized Cys167 residue of the catalytic triad. Its oxidation in the active site cavities of both HaBPS and its mutant T135L is illustrated as dioxide.

This behavior was also observed in the structures of other plant type III PKSs, which were published as MsCHS, GhPS, PCS and OKS corresponding to the PDB numbers 1BI5, 1QLV, 2D3M and 2D52, respectively (Figure V.4). This oxidation of the catalytic cysteine was likewise shown before when the crystal structure of SaBIS1 was solved (Figure V.4) (Raeth, 2007).

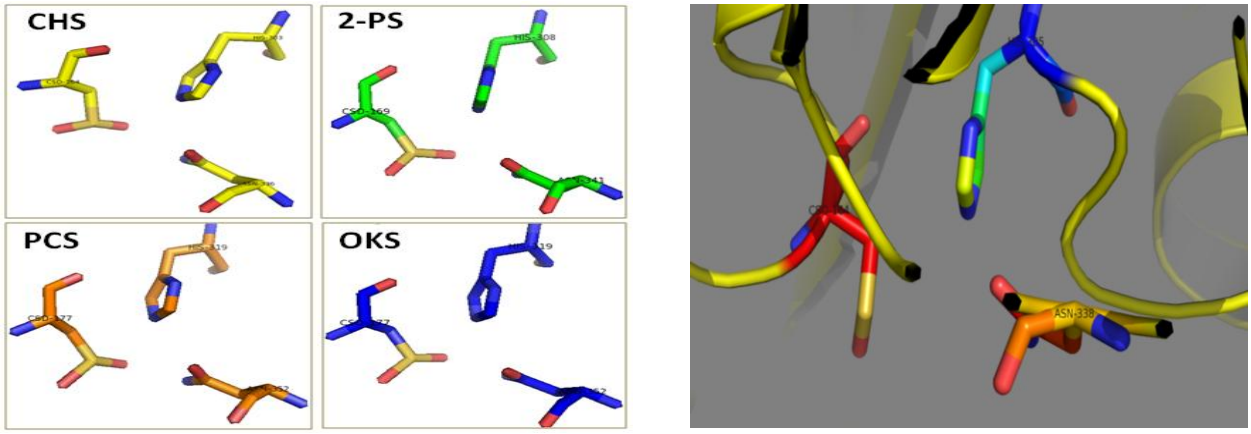


Fig V.4 The oxidized Cys of the catalytic triad was found in many plant type III PKSs (left). The catalytic triad of SaBIS1(right) also shows an oxidation of the Cys164 residue, which is illustrated as monoxide (Raeth, 2007)

1.2.1. The pH change and the use of a reducing agent

The oxidized Cys was thought to be the reason for the unsuccessful soaking or co-crystallization. Firstly, we focused on the pH. Since the purification of the target enzymes was done at basic pH (8.4), protein isolation was then tried under acidic conditions where the pH of the buffers used was adjusted to 6.7 and 7 to protect the active site cysteine from oxidation under basic conditions. Secondly, 4 mM of the reducing agent DTT or 1 mM of the stronger reducing agent TCEP was added to the purification buffers to protect the active site cysteine from oxidation. Not to mention that most of the buffers used in the purification of the already crystallized plant type III PKSs contained 1-2 mM DTT. However, these modifications did not work and the cysteine was always oxidized after new crystallizations screens and structure data analysis.

1.2.2. The use of active sites mutants

The use of the active sites mutants MdBIS3C159A, HaBPSC167A and HsBPSC167S was a trial to open the channel for substrate entrance, in contrast to the oxygen of the oxidized cysteine which may block this entrance (Figure V.5). This trial was done before, the structures of MsCHSC164S with hexanoyl-CoA and MsCHSC164A with malonyl-CoA were solved (Figure V.5). The presence of these two structure complexes in PDB (1CHW and 1CML) may indicate the difficulty of using the wild type enzymes to get the enzyme-substrate/product complexes, for which the reason may be the presence of the unstable cysteine (Ferrer et al., 1999).

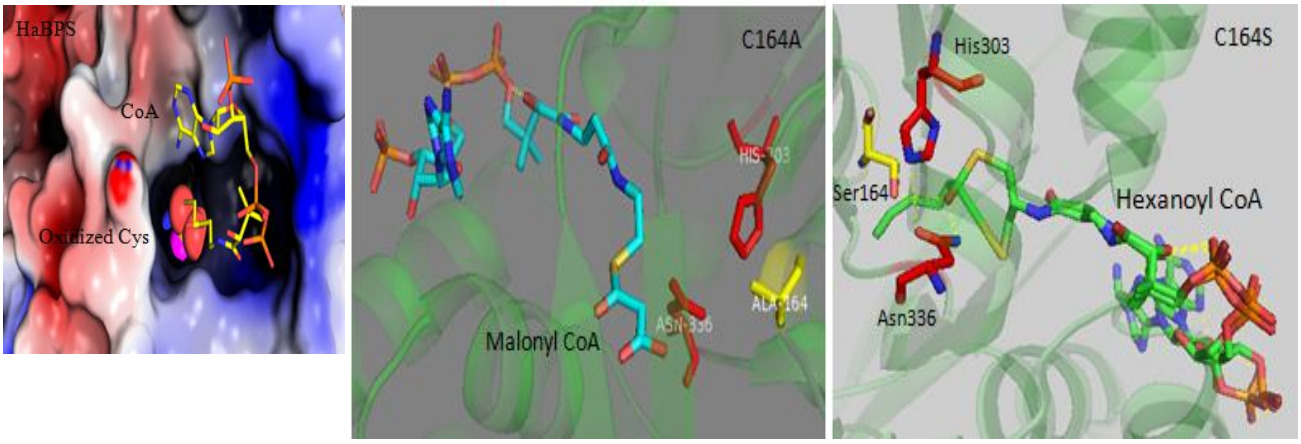


Fig V.5 The oxidized Cys167 in the active site cavity of HaBPS. The oxygen of Cys167, here highlighted as red spheres, may hinders the substrate from entering the active site cavity of HaBPS (left). The absence of active site Cys164 allowed the entrance of malonyl-CoA and hexanoyl-CoA to the active site cavities of MsCHSC164A (middle) and MsCHSC164S (right), respectively.

1.2.3.The presence of an unknown ligand close to Cys167

There was a positive difference density (green mesh) in the active site cavities of both HaBPS and its mutant HaBPST135L, which indicated the presence of an unknown ligand (Figure V.6).

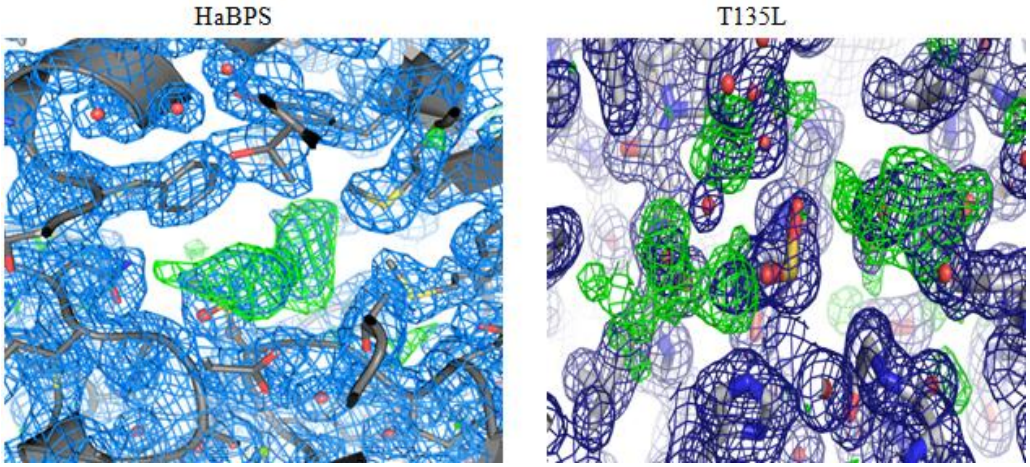


Fig V.6 The presence of the unknown substance (green mesh) in both HaBPS and HaBPST135L, which is in close contact to the catalytic Cys167.

This electron density close to Cys167 could not be modeled to any other compounds that we have used either in the purification process or in the crystallization screens. This manner is also typical for type III PKSs and an example of this is the unintentional presence of malonic acid in the crystal structure of ClCURS (PDB: 3OV2) (Katsuyama et al., 2011). Malonic acid was a component of the crystallization buffer and formed hydrogen bonds with two conserved residues of the active site cavity (His303 and Asn336) (Figure V.7).

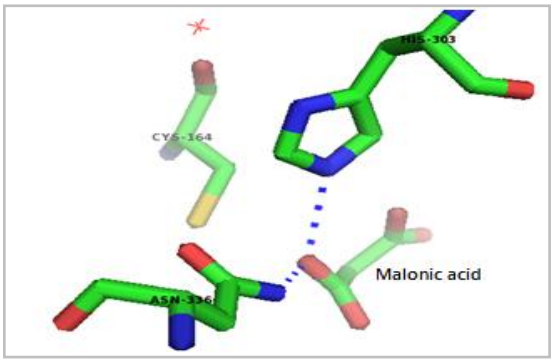


Fig V.7 The unintentional presence of malonic acid in ClCURS. This structure-complex aided in the understanding of the reaction mechanism and the need for a hydrolysis step in curcuminoid production (Katsuyama et al., 2011).

Malonic acid is a mimic to β-keto acid and a mutation was carried out on His303 to study the importance of such an interaction. The mutant H303Q decreased the yield of curcuminoids and, in addition, the kcat of H303Q increased in the use of 3-oxo-5phenyl-4-pentenoic acid more than in the use of cinnamoyldiketide-NAC with feruloyl-CoA, which determined the importance of the hydrolysis step in curcuminoid formation (Katsuyama et al., 2011). In a similar trial, we used benzoic acid for soaking and co-crystallization but were unable to get the complex.

Only HsBPS contained the unoxidized normal sulfhydyl group in the catalytic Cys167 (Figure V.8). Many reasons may explain such a result. One of them may be the development of HsBPS crystals at more acidic pH compared to HaBPS. In addition, the involvement of 1 mM TCEP in the purification buffers has to be mentioned. Therefore, we did not use an active site mutant of this enzyme as in HaBPS and MdBIS3 for soaking or co-crystallization purposes. However, even HsBPS that showed the unoxidized catalytic cysteine for soaking and co-crystallization processes did not form any substrate/product complexes after X-ray measurements and data analysis.

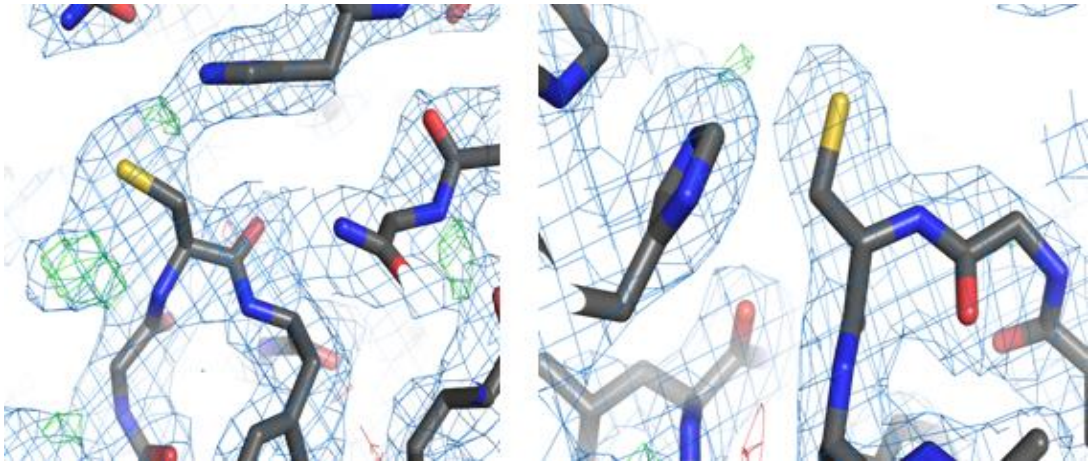


Fig V.8 The crystal structure of HsBPS contained the unoxidized catalytic cysteine. The electron densities around the catalytic cysteine in two monomers of the crystal structure of HsBPS modeled only sulfhydryl groups but not any oxygen atoms.

The recently published structure of MdBIS3 also presented the oxidation of the active site cysteine to sulfinic acid derivatives (–SO2H), which prevented the transfer of substrates onto this active site cysteine (Stewart et al., 2017). Similarly, we always correlated the unsuccesful soaking or co-crystallization to the oxidized cysteine, which was just an assumption because it was not really known if the problem is the cysteine or if something else entered the cavity and blocked the entrance for the substrates. The type III PKS entrance is a solvent-exposed area and many components from the purification or crystallization processes or any other sources can occupy this cavity. This is true as all type III PKSs in PDB contained somehow many components in their active sites cavities. These components, mainly if they are big, can block the way for the substrates to enter, as did the unknown ligand that we saw in the HaBPS and HaBPST135L structures. Actually, if the problem is only the oxidized cystein, there may really be a chance to get an enzyme complex, for example, GhPS contained acetoacetyl-CoA in one of the enzyme complexes after soaking the crystals with 5 mM acetoacetyl-CoA, although the catalytic cysteine was oxidized and 5 mM DTT was used. The oxidation prevented the formation of a covalent acetoacetyl-enzyme complex but allowed the trapping of the acetoacetyl-CoA intermediate inside the enzyme cavity (Figure V.9) (Jez et al., 2000a).

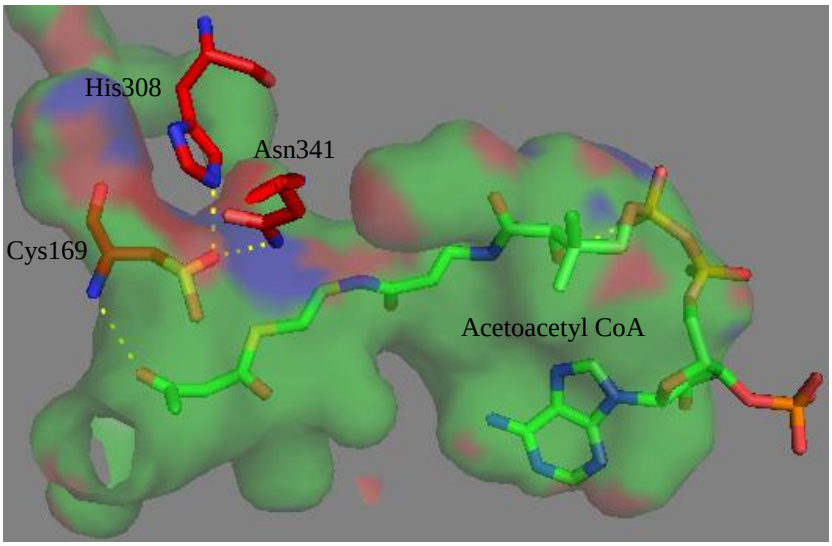


Fig V.9 The GhPS-acetoacetyl-CoA complex (PDB: 1EE0 ; (Jez et al., 2000b, Jez et al., 2000a). Although the catalytic cysteine was oxidized, the enzyme product acetoacetyl-CoA could be trapped inside the active site cavity.

We also used HaBPS and its mutant HaBPST135L in soaking and co-crystallizations processes inspite of the presence of the oxidized catalytic cysteine, but we failed to get any enzyme complexes after X-ray measurement and data analysis. It is not known if the unknown ligand in the active site is the main problem but in one trial we carried out denaturation of HaBPSC167A before going to crystallization using the denaturation agent 8 M urea, which should stimulate components inside the cavity to come out. However, after renaturation by dialysis this protein was not easily refolded back to its correct structure. After gel filtration all the protein aggregated and appeared in the column void volume. In general, refolding of proteins under gradual dialysis is 50% successful and does not work for all proteins.

1.3. Crystal structures of benzoic acid specific type III PKSs

1.3.1. The CoA binding tunnel

The structures of HaBPS and MdBIS3 with CoA were solved at 1.5 Å and 2.5 Å, respectively. Both showed nice orientation of the CoA molecule in the binding tunnel and how this pantetheine arm expands inside the tunnel to push the substrate toward the catalytic Cys. The CoA binding tunnel includes the residues Lys58, Arg61, Lys65 and Gly312 in HaBPS. The only difference from MdBIS3 is the presence of Gly312 instead of Ala305. MsCHS2 contains Ala at this site but both residues involve the same binding pattern with the CoA molecule. Comparing the tunnel of both enzymes with that of MsCHS shows that Arg is the only residue that interacts directly with the CoA molecule and the two Lys residues are just stabilizing the Arg by hydrogen bonding and their other side chains are directing away from the CoA molecule (Figure V.10).

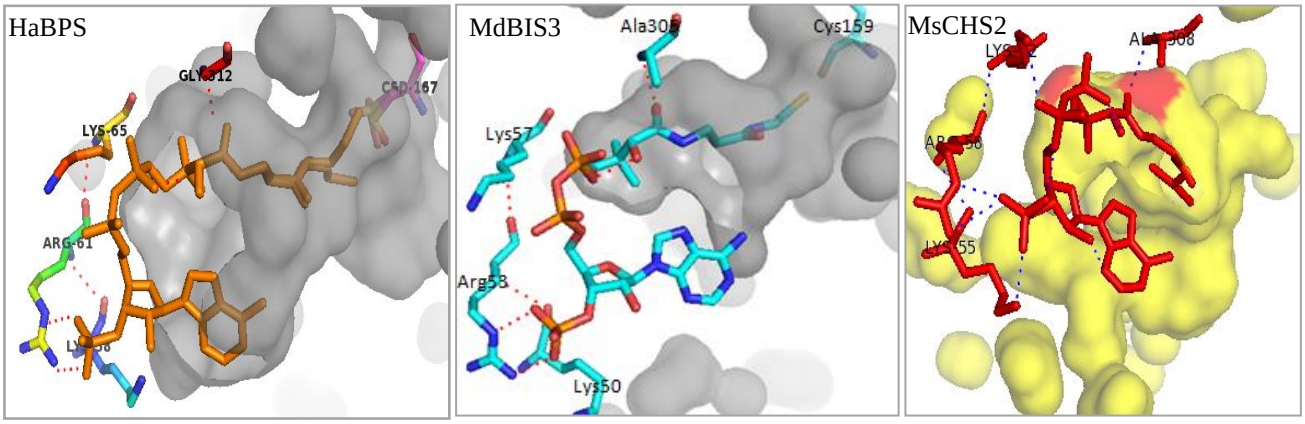
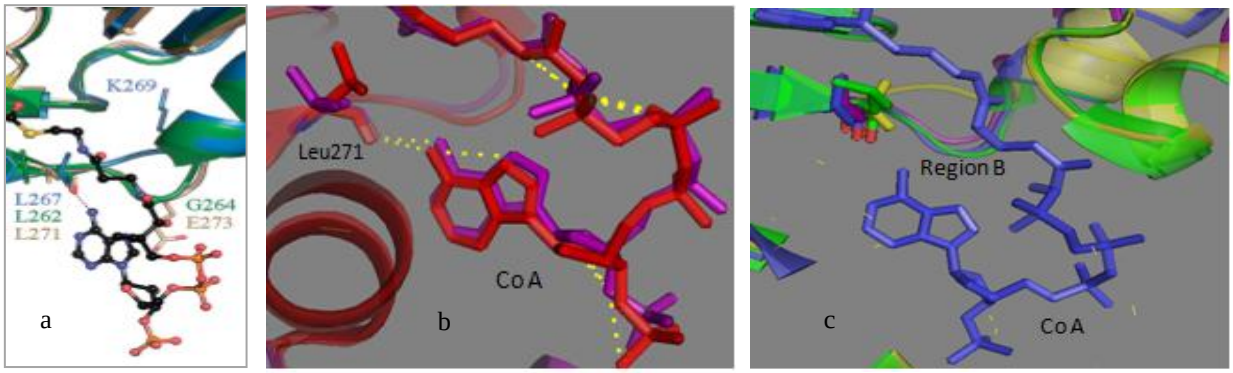


Fig V.10 The CoA-binding tunnel of HaBPS and MdBIS3. In comparison with the CoA-binding tunnel of MsCHS (yellow ; (Ferrer et al., 1999), the different orientation of the two lysine side chains is visible.

A three-residue solvent-exposed loop was observed in the recently published structure of MdBIS3 (Stewart et al., 2017), which is significantly displaced toward the CoA tunnel. It resembles region B in PsSTS mentioned in (I.1.2.3.2.), which did not play any role in the cyclization mechanisms of STS and CHS. This displacement in MdBIS3 resulted in an additional hydrogen bond between the adenine unit of benzoyl-CoA and the backbone carbonyl of Leu262 (Figure V.11a).This additional hydrogen bond was found before in the CoA complex structure of GhPS (Stewart et al., 2017). Our results also demonstrate the clear displacement of this region in both HaBPS and MdBIS3 in comparison to MsCHS2 and PsSTS, but the modeled benzene ring in our MdBIS3-benzoyl-CoA complex did not show the additional hydrogen bond. In contrast, we found this bond in the HaBPS-CoA complex structure (Figure V.11b). This hydrogen bond was found in our HaBPS structure because of the higher resolution of the dataset compared to the published one. We can see in our structures (MdBIS3 and HaBPS) in comparison to the structures of PsSTS and MsCHS2 that the displacement in region B of MdBIS3 is more similar to that in PsSTS than in HaBPS. Interestingly, both enzymes utilize the aldol cyclization mechanism, thus, this region may play a role in the different reaction mechanisms although it is not known yet (Figure V.11c).



V.11 The hydrogen bond between benzoyl-CoA and Leu262 in the published MdBIS3 (Stewart et al., 2017) (a). The structure of our HaBPS-CoA complex (violet) shows hydrogen bond (yellow dotted line) with Leu271, similar to the published structure of MdBIS3 (red) (b). The displacement in the region B loop of our MdBIS3 structure (blue) is more similar to that in PsSTS (green) than in HaBPS (violet). This displacement is significantly different in the loop of MsCHS2 (yellow) in comparison to the other enzymes (c).

1.3.2. The gate keepers Phe 215 and Phe 265 (numbering from MsCHS2)

Phe215 and Phe265 were named as 'gate keeper' because they are located at the active site cavity entrance. Their action by up and down movement of the phenyl rings resembles a gate for allowing and blocking the entrance for compounds between the CoA binding tunnel and the active site cavity (Abe & Morita, 2010). Phe215 participates in the orientation of the terminal carboxylate of malonyl-CoA via van der Waals interactions and perform a non-polar surrounding for carbon dioxide release during the decarboxylation reaction (Jez et al., 2000b). Phe265 is less important than Phe215 but the results of converting *Ruta graveolens* ACS (RgACS) into CHS upon the mutation of its Ala265 into Phe striked the light on this site, too (Lukacin et al., 2001). It proved to participate in the shaping of the active site pocket in addition to its effect on the size and the widening of the entrance of the total cavity (Mori et al., 2013). Phe265 of MsCHS2 is replaced with Tyr in both MdBIS3 (Tyr260) and HaBPS (Tyr269). In HaBPS and MdBIS3, the presence of Tyr with the other gatekeeper residue Phe may also perform the function of the door that allows the entrance of the substrates from the tunnel to the cavity. Interestingly, the orientation of Tyr is similar in both enzymes; however, it is reverse to both residues of gatekeepers in MsCHS2 and PsSTS (Figure V.12).

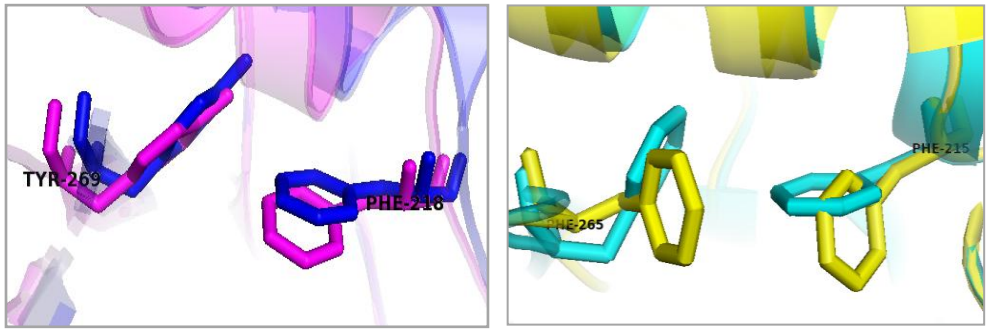


Fig V.12 The gate keepers of benzoic acid-specific type III PKSs. Note the different orientation of the gate keeper residues in HaBPS (purple) and MdBIS3 (blue) in comparison to MsCHS2 (yellow) and PsSTS (cyan) .

The rotameric difference of Phe265 in the published structure of MdBIS3 was correlated to the displacement of the three-residue solvent-exposed loop, which may affect the kinetics of acyl-CoA binding in type III PKS catalysis. The conformation of Tyr260 in MdBIS3 was found to be stabilized by a water-mediated hydrogen bond to the carbonyl of Phe192. It was presumed that this bond is also present in HaBPS, but because of the low resolution diffraction data of the published HaBPS structure it was not possible to see this bond (Stewart et al., 2017). In our structures, we see clearly this water-mediated hydrogen bond in both MdBIS3 and HaBPS (Figure V.13a).

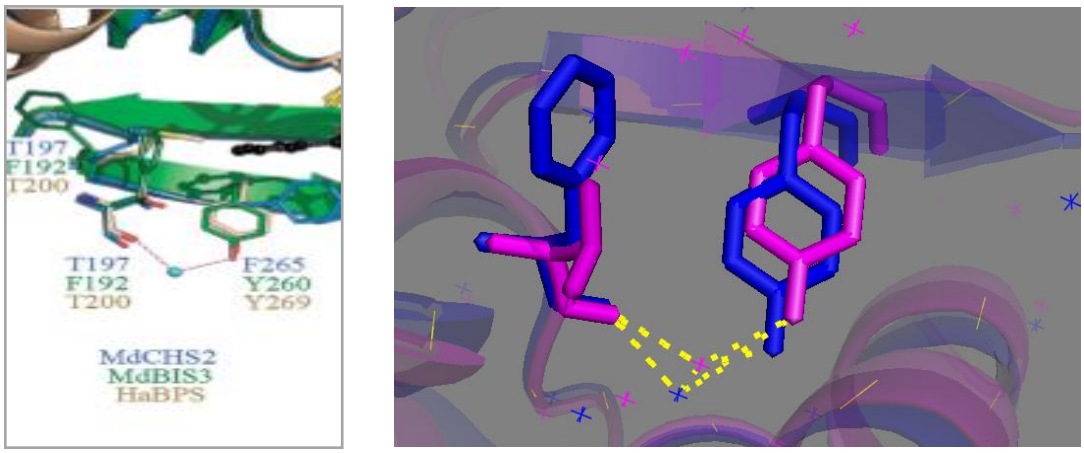


Fig V.13a The water-mediated hydrogen bond in the published MdBIS3 structure (Stewart et al., 2017) ; left). Water-mediated hydrogen bonds between Tyr260 and Phe192 in our MdBIS3 structure (blue) and between Tyr269 and Thr200 in our HaBPS structure (violet).

The rotameric state of Phe265 (MsCHS2) differs consistently between CHS and non-CHS like enzymes (Stewart et al., 2017). Persumely, the conformation of Tyr in our enzymes would meet the need for its hydroxyl group binding which is characteristic for benzoic acid-specific type III PKSs, as we saw this binding also in our resolved structures of HaBPST135L and HsBPS (Figure V.13b). In comparison to other plant type III PKSs, this site was always non polar and the size of the residue payed more attension than its polarity, for example it was replaced by Gly and Leu in OsCUS and CmQNS, respectively, but by Val in CmACS and *Aegl marmelos* (AmQNS).

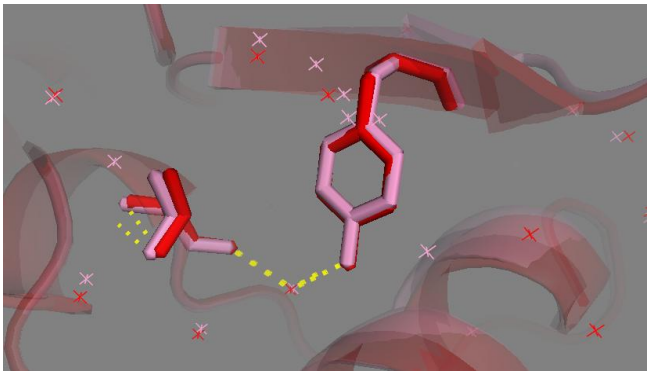


Fig V.13b The water-mediated hydrogen bond between Tyr269 and Thr200 of HsBPS (red) and HaBPST135L (pink).

1.3.3. The substrate binding pocket

The substrates binding pocket of HaBPS enfolds Ser136, Glu195, Thr197, Thr200 and Gly342. These are quite similar to the substrate binding pocket residues of MsCHS2, except for Gly342 which corresponds to Ser338 in MsCHS2 (Figure V.14a). The substrate binding pocket of MdBIS3 enfolds Ser128, Glu187, Thr189, Phe192 and Gly334. These are quite similar to the substrate binding pocket residues of HaBPS, except for Phe192 which corresponds to Thr200 in HaBPS and Thr197 in MsCHS2 (Figure V.14b).

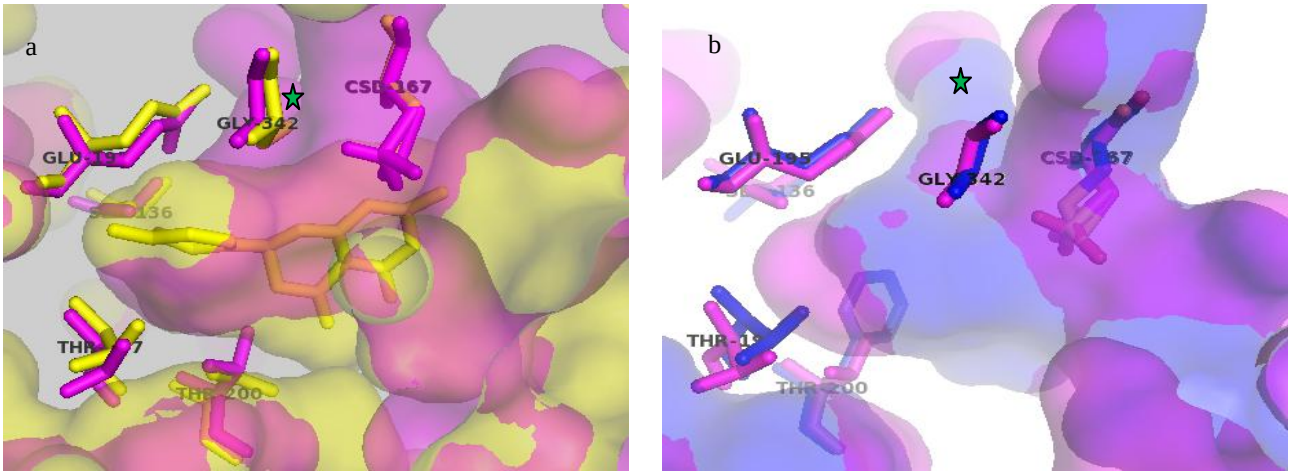


Fig V.14a, b The differences in the substrate binding pockets of MsCHS2 (yellow), HaBPS (violet) and MdBIS3 (blue). Naringinin is shown as yellow sticks representing the coumaroyl-binding pocket of MsCHS2. The green stars represent the new expanded pockets behind Gly342 of HaBPS and Gly334 of MdBIS3.

Ser338 in MsCHS2 forms hydrogen bonds with the neighboring residues Glu192 and Thr132, which participate in water molecule stabilization, thereby triggering the aldol switch mechanism of MsCHS2 (Austin et al., 2004a). This binding is absent in HaBPS because of Gly342 instead of Ser338 and no such hydrogen bond network was found in its crystal structure. This hydrogen bonds network is also impossible in MdBIS3, as Thr132 and Ser338 are replaced with Ala127 and Gly334, respectively, and it was actually not found in the crystal structure of MdBIS3. Mutational experiments proved the importance of Ser338 in polyketide chain length and cyclization because it is close to the catalytic Cys164 and guides the expanding or the proper folding of the linear polyketide intermediate into the active site pocket. For example, the S338I single mutant of MsCHS2 increased the production of methylpyrone in the reaction with acetyl CoA and produced CTAL after the reaction with p-coumaroyl-CoA (Jez et al., 2000a). Ser338 of MsCHS2 is not only absent in HaBPS and MdBIS3 but also absent in other plant type III PKSs. It is replaced by Gly in CmQNS, which is similar to HaBPS and MdBIS3. It is replaced by Ile in GhPS and by Val in AaPCS and AaOKS. It is replaced by Thr in *Rheum palmatum* aleosone synthase (RpALS), which is another plant type III PKS that produces heptaketide intermediate (Abe et al., 2006b). The size of the residue in this site is correlated to the size of the preferred substrate (the side chain of CoA). It is bulk as Ile in GhPS to lock its small preferred substrate acetyl-CoA. It is small as Gly in CmQNS to lock its large preferred substrate N-methylanthraniloyl-CoA. Similarly, the small size of Gly342 in HaBPS and MdBIS3 locks perfectly the benzene ring of its preferred substrate benzoyl-CoA.

When Gly342 was substituted with larger size residues in HaBPS, the single mutants G342F, G342L, G342N and G342Y did not show any enzyme activity and G342A showed less enzyme activity (Klundt, 2008). Only G342S had the same activity like the wild type-enzyme. In addition, any multiple mutations involving this site did not show any enzyme activity, such as A260G/G342S and M267L/Y269F/G342S (Liu et al., 2003). This Gly expands behind the active site cysteines of HaBPS and MdBIS3 to generate novel pockets (Figure 14a, b). Those pockets are specific for MdBIS3 and HaBPS and were not found in the structure of MsCHS2.

Ser133 did not prove to be important for enzymes activity in MsCHS2 because of its distance from the active sites cavity (Austin et al., 2004a). It was replaced with Thr in PsSTS and RpALS and with Ala in GhPS and RpBAS but it was not mentioned to have any effect on substrate specifity changes. Its abscence had an important effect on other type III PKSs. For example, it was replaced with Ala in RgACS and CmACS, which showed to be one of the reasons of owing the ACS larger pocket (Mori et al., 2013). It was also replaced with Ala in AmQNS and CmQNS, which emphasized its effect on the substrate specificity, as the double mutant S132T/A133S accepted p-coumaroyl-CoA and transformed QNS into functional CHS (Resmi et al., 2013). Interestingly, in one of the trials in our group to transform HaBPS into ACS, the produced single mutant S136A did not show any enzyme activity (Klundt, 2008). This site may play an important role in substrates acceptance for benzoic acid-specific typeIII PKSs. Further mutations on this site are not reported for HaBPS or MdBIS3. The site shows different displacement in both enzymes, especially at their hydroxyl group side chains (Figure 14a, b). From the structures, it is not clear if it is important to have a hydrophilic instead of a hydrophobic residue at this site, as we could not see any polar interaction involving this site. Ser136 is also at a distance from the active sites pocket, but its effect on the activity may be due to its close contact to other essential residues in the active site cavity, as with Thr135 and Thr197.

Thr194 of MsCHS2 shows high conservation among plants type III PKSs. It was only replaced with Met in CmQNS. The side chain of Met194 in CmQNS protrudes toward Gly338, which resulted in the absence of the coumaroyl binding pocket of this enzyme. It showed also an influence on the polykeide chain elongation. For example, the single mutant T194M of CmACS resulted in the loss of the tetraketide acridone forming activity and instead in production of the diketide quinolone, so it converted ACS into functional QNS (Mori et al., 2013). Interestingly, the single mutant T197I of HaBPS did not show any enzyme activity. This site covers the lower part of the substrate binding pocket and the prescence of Ile may block this part and hinder the further entrance of the substrate towerd the cavity or the further elongation of the growing polyketides chain. This is also true for the side chain of Thr197 in HaBPS, which takes a part of the active site pocket, in contrast to Thr189 of MdBIS3 which shows high displacement and moves away from the active sites pocket (Figure 14a,b).

Thr197 of MsCHS2 shows high diversity in plant type III PKSs. It is replaced by Leu, Cys, Met, Ala and Gly in GhPS, RpBAS, AaPCS, RpALS and AaOKS respectively. It is replaced by Tyr in OsCUS and CmQNS. Mutational experiments proved the control of this site in polyketide chain length. For example, the single mutant A197T in RpALS completely abolished the heptaketide producing activity and yielded the pentaketide 2, 7-dihydroxy-5-methylchromone (Abe et al., 2006b). In addition, the single mutant T197L of MsCHS produced more methylpyrone than the wild type-enzyme (Jez et al., 2000a). In RpBAS, it did not perform a specific function as the mutant C190T showed the same activity as the wild type-enzyme (Abe et al., 2007b). In HaBPS, the function of Thr200 did not show a clear effect about polyketides chain elongation but it had a direct effect on the activity. Replacing it with bulk residues as Cys and Phe resulted in the production of inactive single mutants (T200C and T200F). Small but hydrophobic residues maintained but decreased the activity as the single mutant T200A was less active than the wild type-enzyme. Double mutants involving this site showed also the same results as T200A / A260G was less active than the wild type-enzyme (Klundt, 2008). This Thr is replaced by Phe192 in MdBIS3, similar to Tyr that is found at this site in OsCUS and CmQNS. This similarity may correlate to similar reaction mechanisms of the three enzymes. MdBIS3 utilizes the condensation of one molecule of malonyl-CoA with 2-hydroxybenzoyl-CoA to form hydoxycoumarin. This single condensation is also performed in CmQNS and OsCUS. The hydroxyl group in Tyr207 in OsCUS was bound to a water molecule which is essential for the hydrolysis step to form β-ketoacid. In addition, the constructed mutant Y207F of OsCUS was not able to produce the main product bisdemethoxycurcumin; otherwise the side product triketide pyrone bisnoryangonin (BNY) was formed (Morita et al., 2010b). In CmQNS, such a water network was not observed and, likewise, we did not see it in our MdBIS3. The orientation of Phe192 in MdBIS3 is more similar to Tyr197 in CmQNS than Tyr207 in OsCUS (Figure 15). This similarity may be important to allow the accommodation of their similar products (4-hydroxycoumarin of MdBIS3 and 4-hydroxy-1-methyl-2-quinolone of CmQNS). Not to mention that both enzymes utilize similar substrates (salicoyl-CoA of MdBIS3 and N-methylanthraniloyl-CoA of CmQNS), which contain similar substitutions of the phenyl ring at the *ortho* position. In addition, similar to MdBIS3, CmQNS also accepts benzoyl-CoA as starter substrate to produce the phenylpyrone.

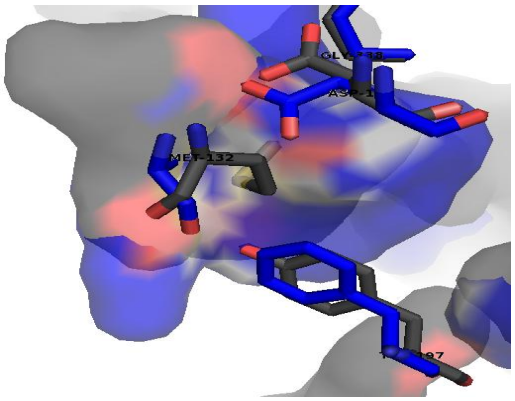


Fig V.15 The substrate-binding pockets of MdBIS3 (blue) and CmQNS (gray). Tyr197 in CmQNS and Phe192 in MdBIS3 have the same orientation. The bulky residues of CmQNS such as Met132 (Ala127 of MdBIS3) decreasesd the size of the pocket in comparison to MdBIS3.

1.3.4. The cyclizations pocket

The cyclizations pocket of HaBPS consists of the residues Thr135, Met140, Phe218, Ile258, Ala260, Tyr269 and Pro379, the numbers and relative positions are the same as in MsCHS2, but Gly256 and Phe265 in MsCHS2 were replaced by Ala260 and Tyr269 in HaBPS, respectively. The cyclization pocket of MdBIS3 consists of the residues Ala127, Met132, Phe210, Val250, Ala252, Tyr261 and Pro371. It differs from HaBPS in two main residues, Ala127 and Val250 corresponding to Thr135 and Ile258, respectively (Figure V.16).

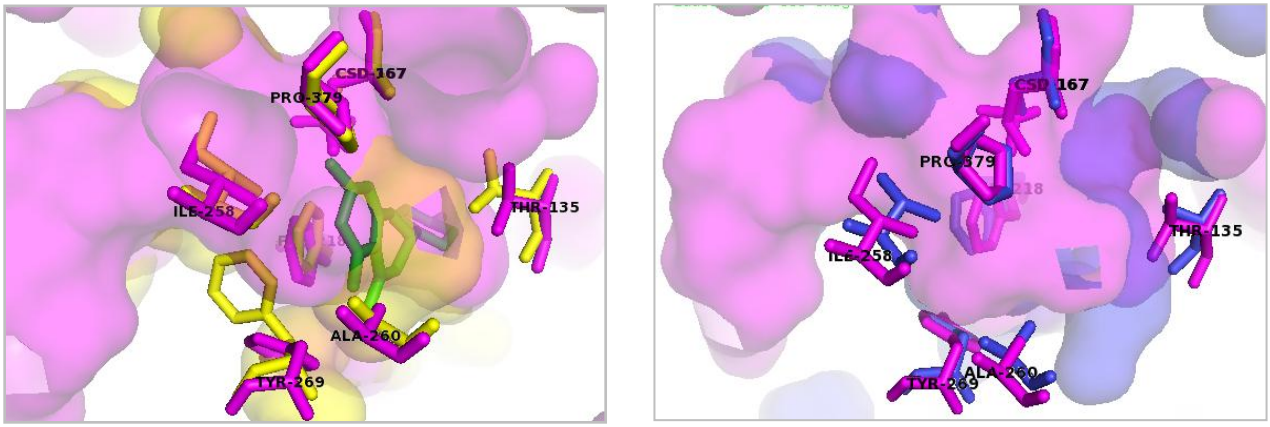


Fig V.16 The differences in the cyclizations pockets of MsCHS2 (yellow), HaBPS (violet) and MdBIS3 (blue). Naringinin is shown as green sticks representing the cyclized aromatic ring after polyketide chain elongation.

Gly256 in MsCHS2 has an impact on the size of the cyclizations pocket and the length of the polyketide chain. It is replaced by Leu in 2-PS, PCS and OKS and, by Met and Ala in OsCUS and CmQNS, respectively. The presence of Ala was proved in the crystal structure of CmQNS to be one of the reasons that resulted in the reduction of the active site cavity (Morita et al., 2010b, Mori et al., 2013). The structure of MsCHS2 and several mutations on Gly256 proved its importance on the length and the cyclization pattern of the polyketide chain. Also, the side chain of this residue is so close to Phe265 and any alteration on Gly256 affects the position of the benzene ring of this residue and thus affects the size and the shape of the active site cavity (Jez et al., 2001). The substitution of Gly256 in MsCHS2 into Ala and Val produced more tetraketide lactone but substitution into Leu and Phe produced more triketide pyrone. Crystal structures of these mutants proved the reduction of the total cavity upon mutation. The side chain of Gly256 extends toward the initiation/elongation/cyclization pocket which has a size of 605 Å, but when it was converted into Ala and Val the size of this pocket decreased to reach 601 Å and 590 Å, respectively. In case of Leu and Phe, the size of the pocket decreased to reach 572 Å and 352 Å, respectively (Jez et al., 2001). Depending on these observations, similar mutations were constructed on HaCHS. The single mutants of HaCHS (G256A and F265Y) did not show any BPS activity. Only when extra mutations were combined with Gly256 and F265, a serious of enzymes were produced and some was found to have more activity toward benzoyl CoA.

For example, F265Y/S338G double mutant and L263M/F265Y/S338G triple mutant preferred benzoyl-CoA over coumaroyl-CoA. In contrast, G256A/L263M/F265Y triple mutant and G256A/L263M/F265Y/S338G quadrant mutant decreased the enzyme activity. Remarkably, similar reverse mutation was carried out on HaBPS, but the mutant M267L/Y269L/G342S did not show any CHS or even BPS activity (Liu et al., 2003). The crystal structure of HaBPS showed a relatively similar big cyclization pocket as in MsCHS2 and CmACS, even when it has Ala260 which is more bulky than Gly256 of MsCHS2 and CmACS (Figure 17a).

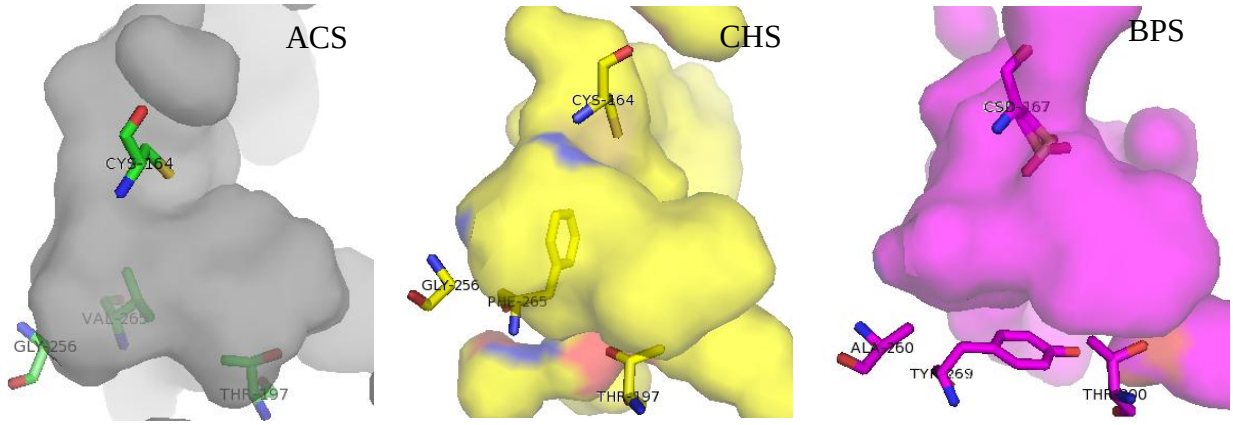


Fig V.17a The similarly big cyclization pockets of CmACS (gray), MsCHS2 (yellow) and HaBPS (violet). The important different residues among the three enzymes are represented as sticks.

CmACS is another plant type III PKSs that utilizes the Claisen cyclization reaction like HaBPS and MsCHS2. It demonstrates a larger cyclization pocket as it has Gly256 as MsCHS2, but in addition the presence of Val instead of Phe265 of MsCHS2 made the pocket even bigger. In case of HaBPS the presence of Tyr269 instead of Phe265 and the specific orientation of its benzene ring made the active site entrance wide as in CmACS (Figure 17b). This may explain the acceptance of N-methylanthraniloyl-CoA by HaBPS but not MsCHS2.

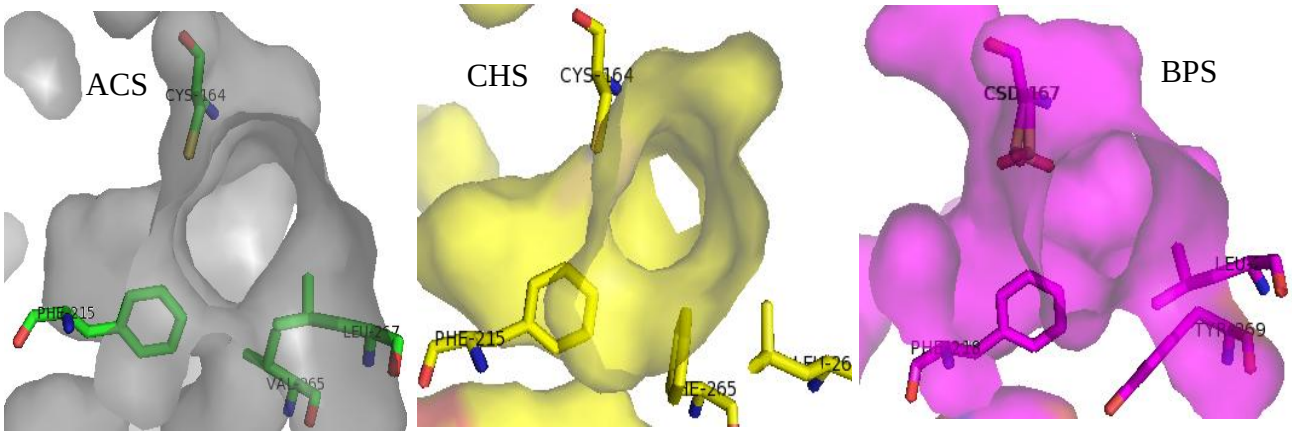


Fig V.17b The wide entrances of CmACS and HaBPS in comparison to the small entrance of MsCHS2. The residues responsible of such variations among the three enzymes are represented as sticks.

Ile254 of MsCHS2, which corresponds to Ile158 of HaBPS, is present in almost all plant type III PKSs. It shows specific replacement into Val250 in MdBIS3. The replacement of Ile254 is not common in plant type III PKSs and it was found only in GhPS where it was replaced by Met259 (Jez et al., 2000a). When Ile258 was replaced by Val in HaBPS, the produced mutant I258V showed similar activity to the wild-type enzyme (Klundt, 2008).

Thr132 of MsCH2 , which corresponds to Thr135 of HaBPS, is the most important known residue of plant type III PKSs, as it was the key that triggered the aldol-switch mechanism of PsSTS (Austin et al., 2004a). It is absent in MdBIS3 and replaced with Ala127. It is also absent in CmACS and AmQNS, which replaces it with Ser, and in CmQNS, which replaces it with Met. It showed an important role in the substrate specificity in addition to its effect on the shape and size of the enzyme pockets (Mori et al., 2013). In one of the trials in our group to transform BPS into BIS, the mutant T135A of HaBPS was constructed which, however, did not show any enzyme activity (Klundt, 2008).

G372FGPG loop of MsCHS2, which corresponds to G375FGPG loop of PsSTS, is known as conserved loop in plant type III PKSs. It was found in HaBPS and MdBIS3 but with one substitution. The loop comprises G376LGPG in HaBPS but G368IGPG in MdBIS3. The flexibility of this loop proved to be important in the diversity of the compounds produced by STS and CHS (Suh et al., 2000). As the GPG part showed conservation in all plant type III PKSs, the first two residues showed variation among them. For example, in CmQNS and OsCUS two other loops were found, which are A372IGPG and A383FGPG, respectively. In the structures of HaBPS and MdBIS3, there was a clear displacement in the loop residues, especially at the substitution sites of Leu377 in HaBPS and Ile369 in MdBIS3. This is contrary to the G372FGPG loop of MsCHS2 and the G375FGPG loop of PsSTS, in which their loops are similar and did not show an obvious displacement (Figure V.18). Leu377 was subjected to serious of mutations in HaBPS, the mutants L377G and L377S decreased significantly the activity, but the mutants L377T and L377F produced inactive enzymes. The hydrophobic side chain of Phe and the bigger size of Thr in comparison to the smaller residues Gly and Ser may hinder the elongation of the growing polyketide chain because of the reduction in the size of the lined pocket (Klundt, 2008).

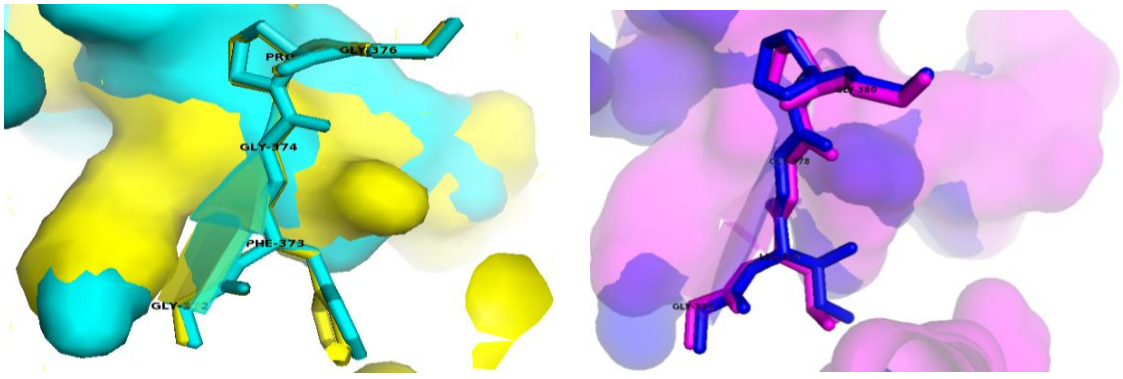


Fig V.18 The displacement of the residues in the G368IGPG loop (blue) of MdBIS3 and the G376LGPG loop of HaBPS (purple). In comparison to Phe373 of the MsCHS2 loop (yellow) and Phe376 of the PsSTS loop (cyan), the prescence of Leu377 of the HaBPS loop and Ile369 of the MdBIS3 loop allows their pockets to expand.

The active sites pocket of HaBPS shows a totally different shape and geometry from MsCHS2, especially at the areas involved in the previously mentioned mutations sites. Those mutations depended mainly on amino acids sequence differences, but the crystal structures showed that even similar residues may have different orientations and may affect the geometry of the active site cavity. An example of this is the different orientation of Leu in the cyclizations pocket (Leu271 in HaBPS, Leu267 in MsCHS2) (Figure V.19). Signicant changes in the active sites pockets of MsCHS2, HaBPS and MdBIS3 do not arise from the known conserved residues of plant type III PKSs, but actually from their neighboring residues that show high diversity especially at specific regions in the HaBPS and MdBIS3 structures. For example, the conserved residues Thr197 of MsCHS2 and Thr200 of HaBPS, which corresponds to Phe192 of MdBIS3, protude in the same direction toward their active site pockets. In contrast, Phe192 has the same orientation like the neighboring residues Leu263 of MsCH2 and Met267 of HaBPS. Phe258 of MdBIS3, which corresponds to Leu263 of MsCH2 and Met267 of HaBPS, moves away from the active site pocket (Figure.19). The presence of the bulk residues Phe192 of MdBIS3 and Met267 of HaBPS explains why both enzymes prefer the hydrophobic substrate benzoyl-CoA, which contains the small size hydrophobic benzene ring. In addition, both residues clash sterically with the carbonyl unit of naringenin when a CHS–naringenin complex is superimposed with MdBIS3 and HaBPS structures. This explains why both enzymes can not utilize the bigger substrate 4-coumaroyl-CoA like MsCHS2.

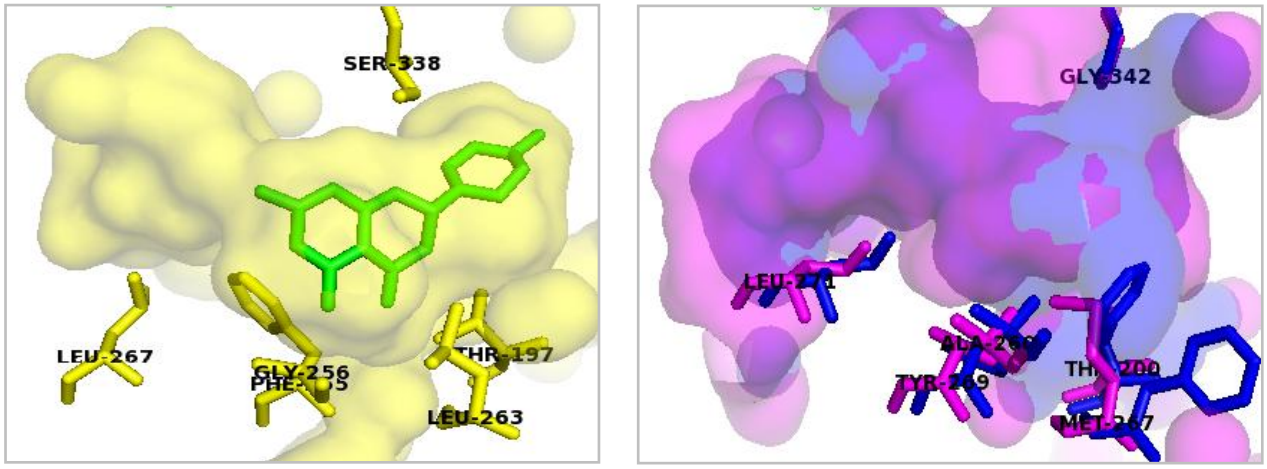


Fig V.19 The important different residues in the active site cavities of HaBPS (violet), MdBIS3 (blue) and MsCHS2 (yellow). These residues were mainly used for mutational experiments. Phe192 of MdBIS3 and Met267 of HaBPS clash sterically with the carbonyl unit of naringenin (green).

Collection of all different residues surrounding the initiation/elongation pockets of both enzymes strikes the light on three main different residues of MdBIS3, i.e. Val191, Phe192, and Phe259 which are replaced with Met199, Thr200, and Met267 in HaBPS respectively (Figure V.20). The single site mutants of HaBPS, M199V and M267F, were reported to have the same activity like the wild-type enzyme but the single mutant T200F and the double mutant M199V / T200F did not show any activity (Klundt, 2008). A triple mutant involving these three sites is not reported.

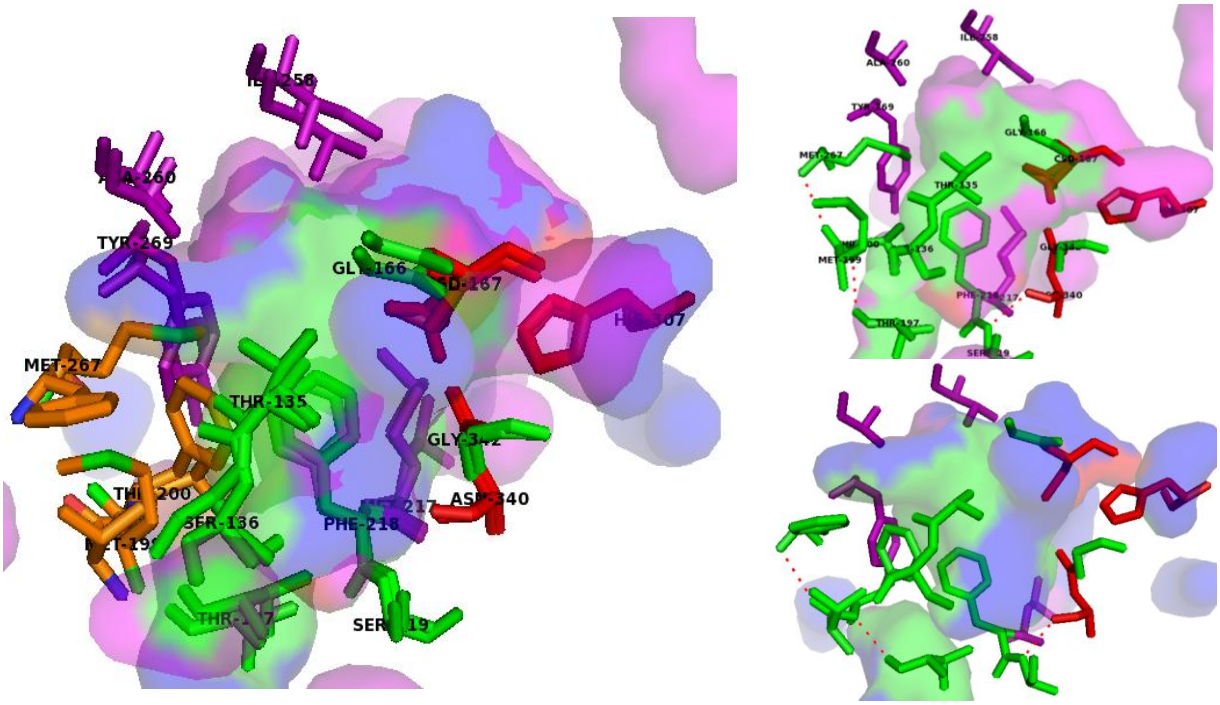


Fig V.20 The different orientation of the initiation pocket residues (violet) and the elongation pocket residues (green) between HaBPS and MdBIS3. The catalytic triad residues (red) and the main different residues (orange) are highlighted. On the right, the polar interaction (red dotted line) between the initiation/elongation pocket residues of MdBIS3 (blue) and HaBPS (purple) is depicted.

The active sites cavities of both HaBPS and MdBIS3 are totally different from each other and varied completely from the ones of MsCHS2 and PsSTS in their geometry. They show displacement at many loops surrounding their main active site pockets. There are common similarities between them and CmQNS as they show G256A/S338G substitutions in comparison to MsCHS2.

Specifically, MdBIS3 and CmQNS demonstrated substitution of Thr197 into Tyr (Figure V.21). These similar features grouped both enzymes so close to CmQNS in the phylogenetic tree (Mori et al., 2013).

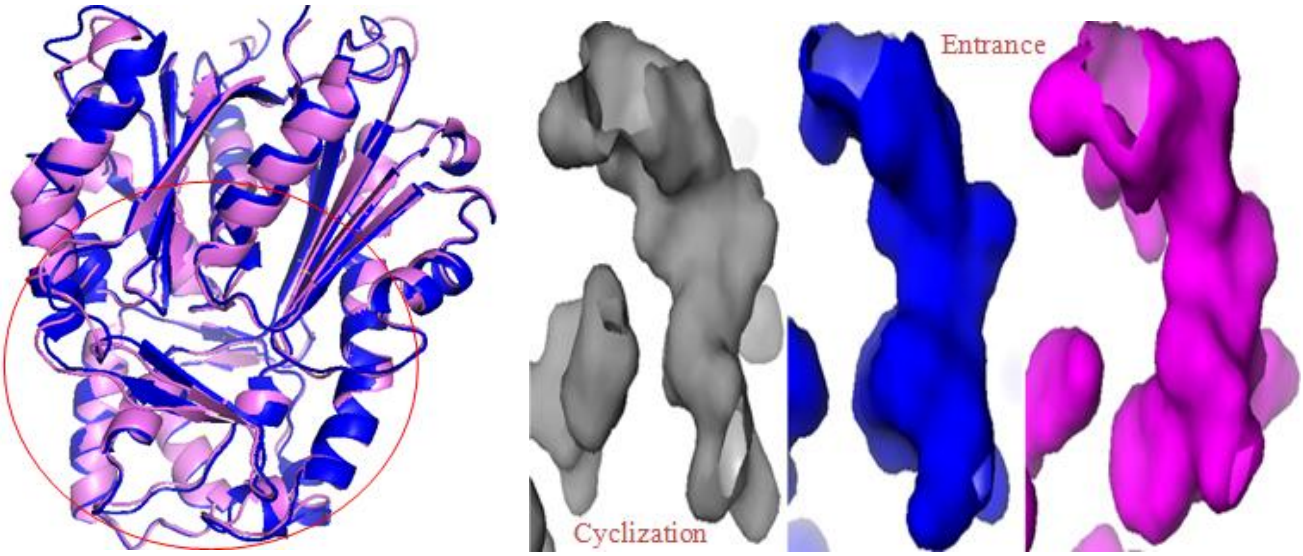


Fig V.21 The similar active site pockets of HaBPS (violet), MdBIS3 (blue) and CmQNS (gray). On the left, the marked red circle demonstrates the site of the active site cavities, which shows displacement in many loops of HaBPS and MdBIS3. On the right, the entrance of HaBPS (violet) is big and similar to CmQNS (gray), which explains why both enzymes accept the big substrate *N*-methylanthraniloyl-CoA in contrast to MdBIS3 (blue). The cyclization pocket of CmQNS is not as big as that of MdBIS3 and HaBPS, which explains why both enzymes form a tetraketide rather than monoketide like CmQNS.

1.3.5.The new hidden pocket of benzoic acid-specific type III PKSs

The structure of HaBPST135L was solved at 1.4 Å, showing the proper geometry of the residues surrounding the mutation site by formation of a new expanded pocket (Figure V.22a). The orientation of Leu135 was supposed to be the reason for the low activity of the mutant with 3-hydroxybenzoyl-CoA. The growing 3-hydroxybenzoyl triketide functions like the absent Thr135 and forms hydrogen bond with the neighboring Gly166, which results in the trapping the produced polyketide chain inside the new pocket (Figure V.22b) (Klundt et al., 2009).

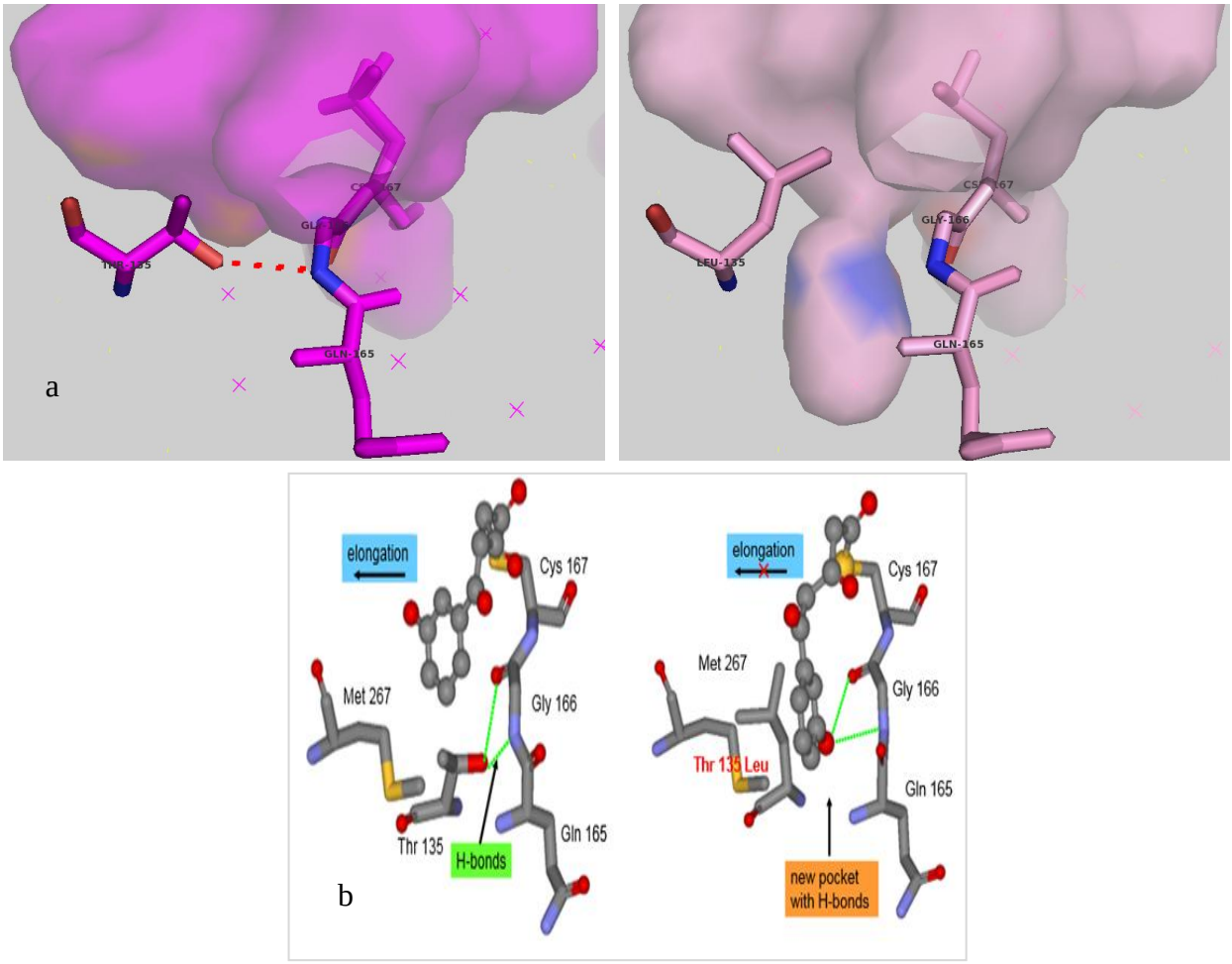


Fig V.22 The different orientations of Thr135 in HaBPS (purple) and Leu135 in HaBPST135L (pink). Thr135 forms a hydrogen bond (red dotted line) with the neighboring residue Gly166 (a). The trapping of 3-hydroxybenzoyl-CoA in the newly formed pocket blocks the elongation reaction on this substrate (b) (Klundt et al., 2009).

The size and shape of the active sites pocket is strongly correlated to the number of condensations with malonyl-CoA and the length of the polyketide chain. After comparison of the active site pockets of wild-type HaBPS, which condenses three molecules of malonyl-CoA, and the mutant HaBPST135L, which condenses only two molecules of malonyl CoA, we can see a clear decrease in the active site pocket size of HaBPST135L. There is a new geometry surrounding L135 which is shown clearly by the slight displacement in Gln165 (Figure V.23). The entrance of the pocket is small and a phenyl moity cannot be locked inside it. The decreased size of this pocket obviously limits the chain extension to only two molecules of malonyl-CoA addition.

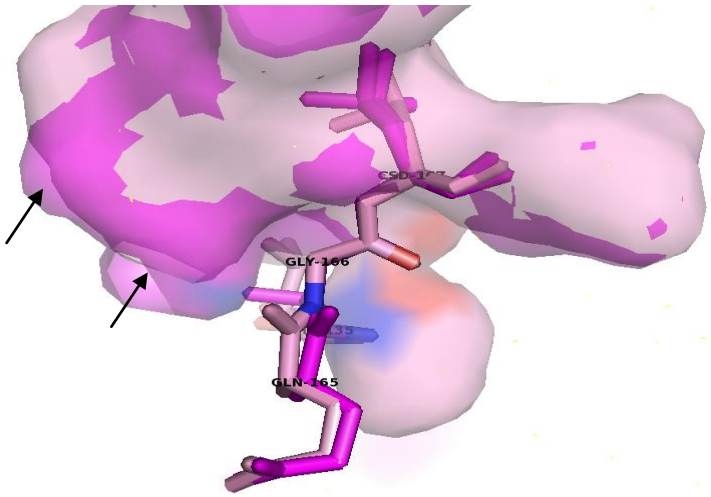


Fig V.23 The different pocket sizes of HaBPS and HaBPST135L. The size of the pocket surrounding Leu135 in HaT135L is smaller than the pocket surrounding Thr135 in HaBPS (indicated with arrows). This explains the production of a triketide chain by the mutant instead of a tetraketide chain by the wild-type enzyme.

Interestingly, Ala127 in MdBIS3 does not participate in any polar interaction, as Thr135 of HaBPS. Therefore, it acts exactly as Leu135 in HaBPST135L to create a bigger new pocket. In the recently published work about MdBIS3, a hidden pocket was discovered in the active site cavityies of both MdBIS3 and HaBPS. The side chains of Ala127 in MdBIS3 and Thr135 in HaBPS line the sides of this novel pocket. Actually, this pocket was found in the structures of our HaBPS and MdBIS3.This pocket is located exactly as the one that we found in the HaBPST135L structure, which was also discovered previously in our research group by homology modeling of HaBPS and the mutant HaBPST135L. This novel pocket was hypothesized to favor the production of phenylpyrone over phlorobenzophenone (Klundt et al., 2009). This new pocket is smaller in the solved crystal structures of the two BPSs (HaBPS and HsBPS) because of the hydrogen bond bridge of Thr135, but larger in MdBIS3 and the T135L mutant because of the abscenc of Thr135 (Figure V.24a).

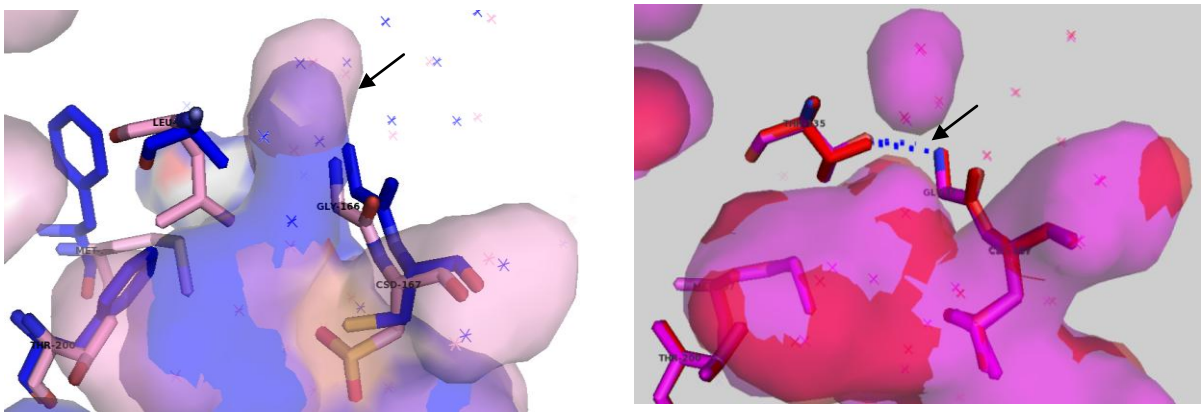


Fig V.24a The new pocket (marked with an arrow) of benzoic acid-specific type III PKSs. There are similarly big pockets in MdBIS3 (blue) and HaBPST135L (pink). The size of the new pockets of HaBPS (violet) and HsBPS (red) is small. The presence of Thr135 hydrogen bonds (blue dotted line) is the reason for the smaller pockets.

Comparing our structures of HaBPS and MdBIS3 demonstrates the bigger new pocket in MdBIS3 than in HaBPS. This is bcause of the presence of more polar residues around this pocket in HaBPS than MdBIS3. These are mainly Thr135 and Ser343 of HaBPS, which correspond to Ala127 and Ala336 in MdBIS3. Thr135 and Ser343 of HaBPS are involved in many hydrogen bond networks, which minimize its new pocket size. These residues are localized in two main areas in both enzymes, which show high variation between them. They are V130VFATTSGF (a1) and G342SACVMF (a2) in HaBPS, corresponding to L122IFCTASCV and G335APTVHF of MdBIS3, respectively (Figure V.24b). The recent publication about the structure of MdBIS3 also mentioned, that sequence differences between the residues lining the walls of the novel pockets in BPS (T135,G342SAC) and BIS (A127, G335APT/S) are good targets for mutagenesis and a useful starting point for further exploration of the cyclization specificity. E195 of HaBPS (E192 of MsCHS2) locates between the previously mentioned regions a1 and a2. In addition, the suggested water-hydrogen bonds network involves another area (N164QGF of HaBPS and E156AGY of MdBIS3) (a3), which includes the important residue G166 of HaBPS. This area did not show high variation or notable displacement in comparison to MdBIS3. In contrast, downwards and close to E195 of HaBPS there is another important area (E195NTAMT of HaBPS and E187ITTVF of MdBIS3) (a4). Two main different residues in a4 (Thr197 and Thr200 of HaBPS corresponding to Thr189 and Phe192 of MdBIS3) show displacement among HaBPS and MdBIS3. From the structure, we see clearly that those two residues sterically decreased the size of the MdBIS3 cavity compared to that of HaBPS.

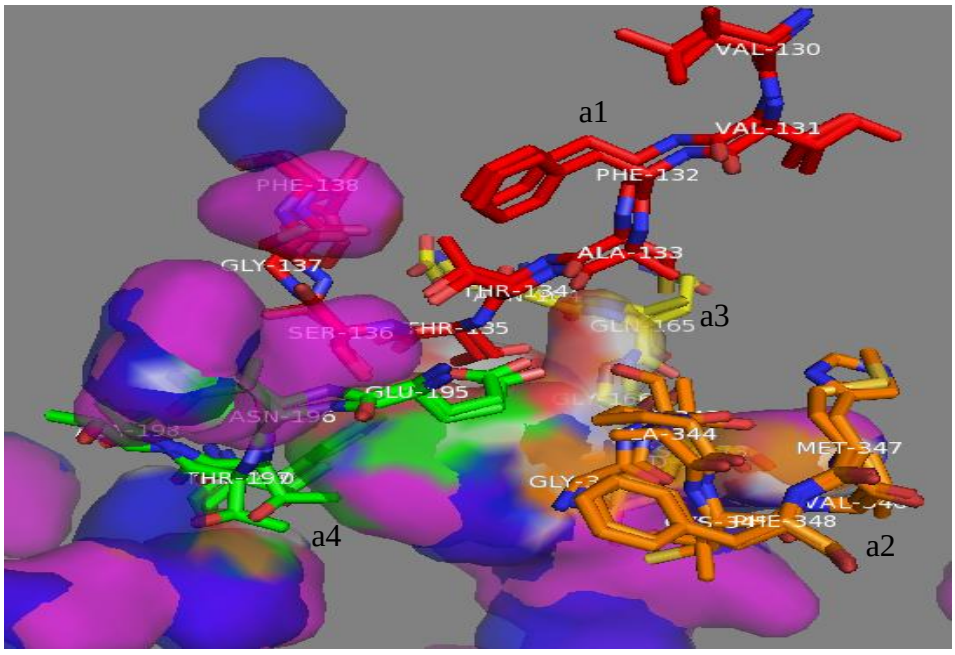


Fig V.24b The four different areas surrounding the new pockets of HaBPS (violet) and MfBIS3 (blue). a1 (red), a2 (orange), a3 (yellow) and a4 (green).

1.3.6. Aldol/Claisen cyclizations mechanisms and the electronic/steric effect

The new publication (Stewart et al., 2017) also assumes, that the novel pocket of MdBIS3 contains ordered waters, which could function similar to the nucleophilic water that is essential for the aldol switch of stilbene synthases, but the low resolution structure of HaBPS prohibited further analysis. Interestingly, we see clearly in our structures of HaBPS and MdBIS3 different hydrogen bond network with one interesting water molecule. This water is located exactly as the water molecule found in MsCHS2, PsSTS, AhSTS and even the 18xCHS (Figure V.25). The only difference from our HaBPS and MdBIS3 is the location and orientation of Thr135 of HaBPS and Ala127 of MdBIS3 (corresponding to Thr132 of MsCHS2). In HaBPS, Thr135 directs toward Gly166, which assists their hydrogen binding. Similar hydrogen binding of Thr132 and Gly163 was found only in MsCH2. In contrast, Thr132 of AhSTS and Thr135 of PsSTS direct away and do not involve any binding with their corresponding Gly163 and Gly166, respectively. In MdBIS3, there is no contact with its Gly158, as the enzyme has Ala127 instead of Thr. This is similar to HaBPST135L, which has Leu135 instead of Thr; the only difference is the reduction of the active site cavity of HaBPST135L because of the bulky Leu. This reduction resulted in triketide chain elongation and the formation of phenylpyrone. These variations are still related to the presence or the orientation of Thr132 and Ser338 of MsCHS2, which are the responsible residues for the aldol-switch mechanism in STS. These residues are still the critical residues in our enzymes also in addition to their interaction with Gly163 of MsCHS2, which resulted in the expanding of a small hidden pocket and might strongly relate to the aldol/Claisen cyclization mechanisms (Figure V.25).

The orientation of Thr132 and Gly163 of MsCHS2 is the same as Thr135 and Gly166 of HaBPS. Both enzymes utilize the Claisen cyclizations reaction. Both residues are directed toward each other using one specific hydrogen bond. This bond will in turn affect the close hydrogen bonds network that involves the important water molecule. The only difference is the absence of Ser338 of MsCHS2 (Gly342 of HaBPS), which resulted in the formation of the expanded new pocket. In MdBIS3, there is Ala127 instead of Thr135 in HaBPS, which resulted in a bigger new pocket. Ala127 of MdBIS3 does not involve any binding with Gly158, this is similar to AhSTS, PsSTS and 18xCHS, which all utilize aldol cyclization reaction. Another two plants type III PKSs were analysed in their structures at the positions of the previously mentioned two residues. CmACS (Ser321 and Gly163) showed the same orientations of MsCHS2 and HaBPS and this enzyme also utilizes the Claisen condensation. CmQNS (Met132 and Gly163) showed the same orientation of PsSTS and MdBIS3 and this enzyme also utilizes aldol condensation (Figure V.25).

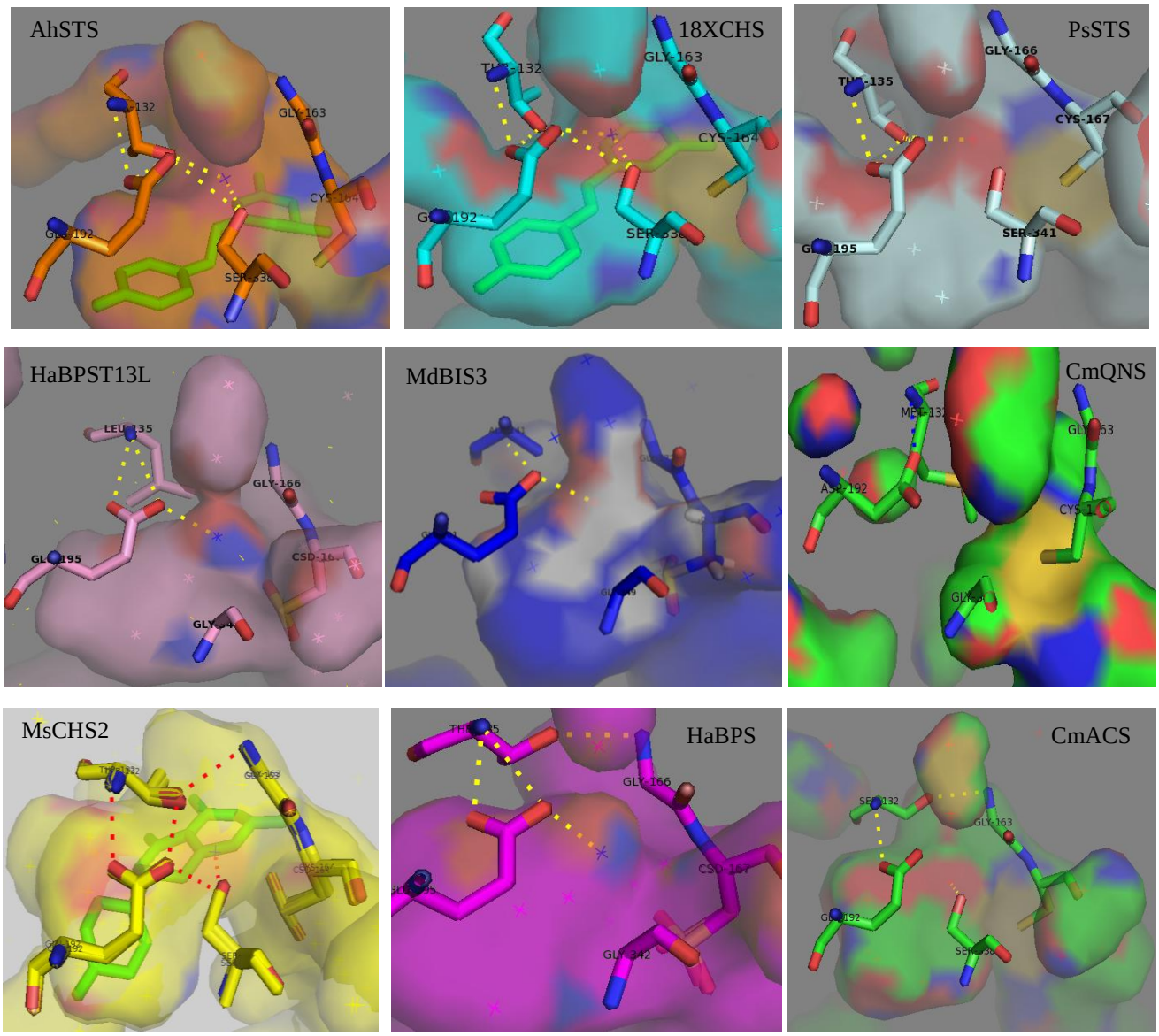


Fig V.25 The different orientations of Thr132 and Gly166 in MsCH2 and their corresponding residues in other plant type III PKSs. MdBIS3 utilizes the aldol cyclization reaction and shows no hydrogen bond between these residues similar to other enzymes utilizing the same mechanism (AhSTS, PsSTS and CmQNS). HaBPS utilizes the Claisen cyclization reaction and shows hydrogen bonding between the previously mentioned residues, similar to other enzymes utilizing the same mechanism (MsCHS2 and CmACS). The water molecule responsible for the aldol switch in PsSTS was found in the same location in other type III PKSs but in different hydrogen bond networks (yellow dotted lines). This water is close to the pocket where the cyclization takes place. Resveratrol in AhSTS and18xCHS and naringenin in MsCHS2 are shown as green sticks.

For the hidden pocket of benzoic acid-specific type III PKS, the bulk size of the surrounding residues affects its formation, Also, the orientation or the absence of Thr132 in relation to the presence of Ser338 of MsCHS2 and their hydrogen bonds network with the surrounding residues that stabilize a localized water molecule in a constant place in most of plant type III PKSs which is close to the catalytic Cys164 and the cyclizations pocket indicate that the steric factor together with the electronic factor are the trigger for utilizing either aldol or Claisen cyclization mechanisms. Benzoic acid-specific type III PKSs demonstrate totally different active sites pockets compared to those of MsCHS2 and PsSTS. Further mutagenesis works are needed to study the steric/electronic effects especially at residues surrounding Thr132 and Gly163 of MsCHS2. Actually in the literature, there is one study on AhSTS that ensures that the steric factor is more important as the electronic factor, which is known as aldol-switch of PsSTS. This study demonstrated two hydrophobic residues Gly255 and Met98 that play an important effect in CHS-STS functions variation. The structure of AhSTS (peanut) demonstrated also three main different areas, a1 (133–136), a2 (248–256) and a3 (265–269) after comparison with the structure of MsCHS2. As a part of a1, Thr132 has the same position that was found in PsSTS and it also stabilizes a water molecule by contribution into hydrogen bond network with Glue192 and Ser338 (Figure V.32). In contrast to the previous mentioned study about the crystal structure of PsSTS, this one focused more on the steric effect as an important factor that influences the cyclizations mechanism of the enzyme. By structures comparison of AhSTS with MsCHS2, the side chain of its Met98 hinders the structure displacement in a1 by the bulkiness of its side chain and it is still not clear if the substitution of this residue alone or with other additional substitutions will have effect on the functional conversion between CHS and STS. More interesting is the substitution of Gly255 with bulkier residues, which resulted in the conversion of STS into functional CHS. It seems that there are more than one factor effecting the cyclization pattern of type III PKSs (Figure V.26) (Shomura et al., 2005). Interstingly, those two residues Met98 and Gly255 of AhSTS are close to Thr132 and Gly153. The substitutions of Met98 and Gly255 with other residues surly disrupted the orientations of Thr132 and Gly153 and disrupted also their close water hydrogen bonds network.

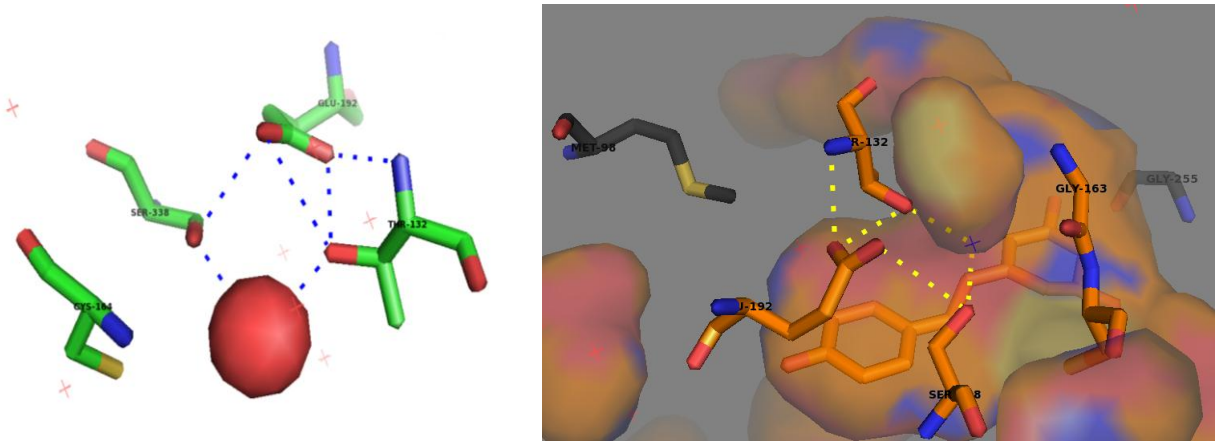


Fig V.26 The hydrogen bond network in peanut STS is similar to that in PsSTS (left). On the right, the orientation of Met98 and Gly255 (black) in AhSTS is depicted.

Met98 is replaced with Cys93 in MdBIS3 but it is the same residue of MsCHS2 (Leu98) and HaBPS (Leu101). The closeness of this residue to Thr132 of MsCHS2 should effect its orientation. Gly255 is so close to the cyclizations pocket and the part of the product that utilizes aldol or Claisen reaction. It is replaced with Asp255 in MsCHS2, which changes the size of the pocket that locks the two hydroxyl groups of the second cyclized ring of naringenin. It is replaced with Thr259 and Val251 of HaBPS and MdBIS3, respectively (Figure V.27). This replacement shows a clear change in the size of the pocket that should lock, similarly to naringenin of MsCH2, the two hydroxyl groups of the second cyclized ring of biphenyl and phlorobenzophenone. The structure of enzyme-product complexes are needed in the future for further structure analysis and the interpretation of the different cyclization mechanisms.

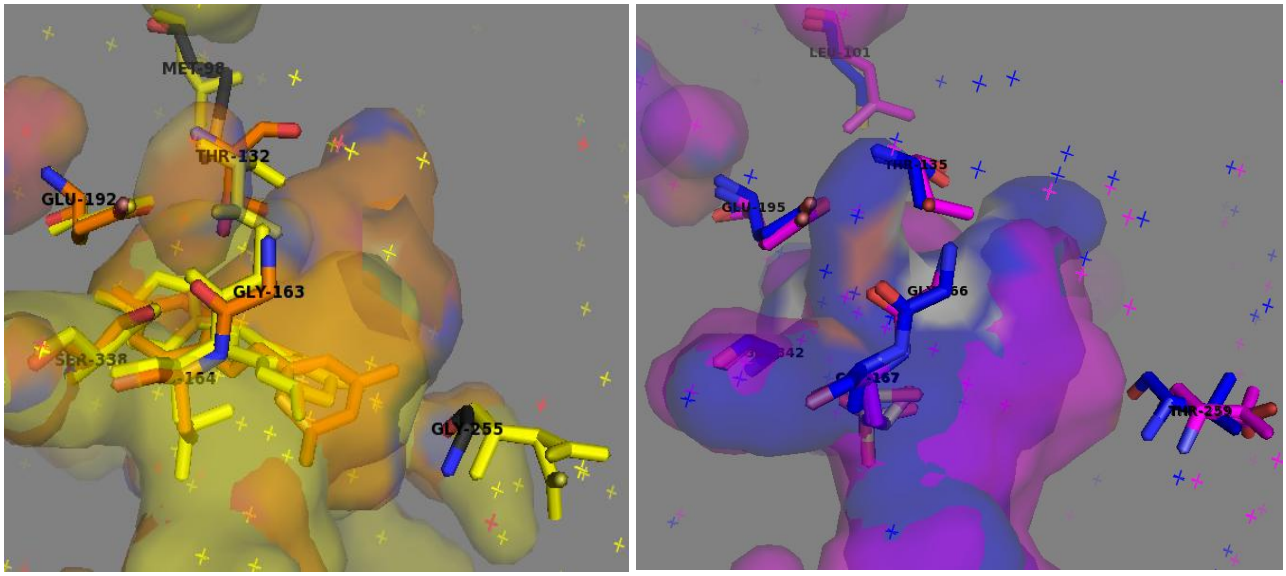


Fig V.27 The orientation of Met98 and Gly255 (black) in AhSTS (orange). In comparison to MsCHS2 (yellow), HaBPS (violet) and MdBIS3 (blue), Met98 is so close to Thr132 and Gly255 that it results in different shapes and sizes of the pocket close to it in all the enzymes.

The importance of the steric effect on the cyclization mechanisms was also reported for other type III PKSs because of the absence of the water molecule that is responsible for the aldol mechanism in PsSTS. An example is the aldol-Claisen cyclisations mechanisms alteration after mutagenesis in AaPCS and AaOKS. In AaPCS, an intrinsic character was found when M207 (Phe192 of MdBIS3 and Thr200 of HaBPS) was substituted by Gly. The produced mutant utilized aldol rather than Claisen cyclization mechanism like the wild-type enzyme. The structure analysis supposed that the alteration of the reaction mechanism of AaPCS C1/C6 Claisen type cyclization into C10/C15 and C12/C7 aldol-type cyclization of the M207G mutant is because of the steric effect rather than the electronic effect (Figure V.28) (Morita et al., 2007).

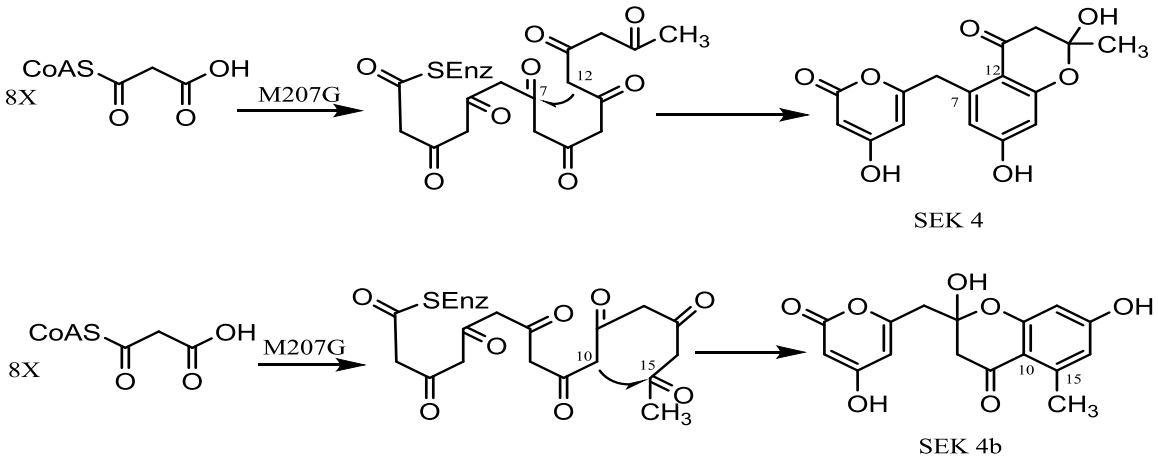


Fig V.28 The aldol cyclization mechanism of the M207G mutant of PCS.

The electronic effect was explained previously as aldol switch of PsSTS because of the presence of a hydrogen bonds network close to Ser341 (Ser338 in MsCHS2). Such network was not found in the structures of AaPCS or its mutant M207G. Similar to MdBIS3, this Ser was absent (Gly334) and replaced with the hydrophobic residue Val350 in AaPCS and its mutant. It was suggested that the location of residue 338 is the most important factor as it is so close to the catalytic Cys that is attaches to the growing polyketide chain. Thus, it performs a steric guidance for the linear intermediate to be extended toward other hidden pockets of the active site cavity (Figure V.24) (Morita et al., 2007). Moreover, the impact of the steric effect on the reaction mechanism was shown when the Claisen cyclization of AaPCS changed in the triple mutant (F80A/Y82A/M207G) into the aldol condensation pattern to release CO_2 and to form the aromatic naphthalene ring (Figure V.29) (Abe et al., 2007a).

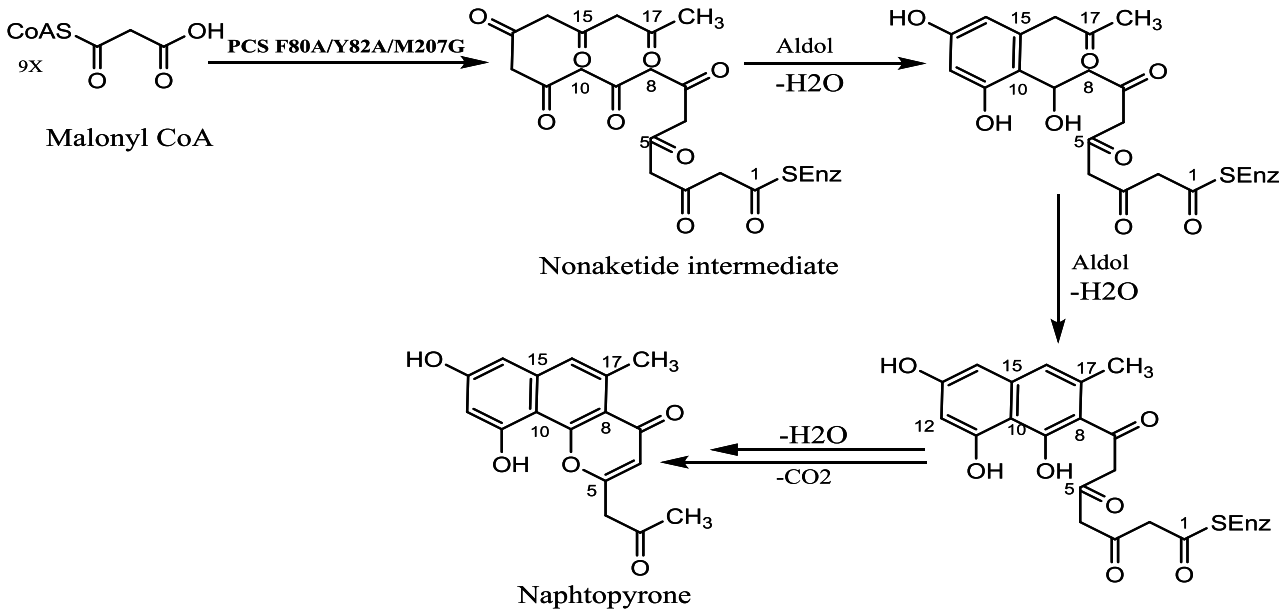


Fig V.29 The aldol cyclization mechanism of the F80A/Y82A/M207G mutant of PCS.

Even more interesting, the mutant G207M of AaOKS (that has PCS activity) shows intrinsic mechanism than the natural AaPCS. Normaly, AaPCS resulted in 5,7-dihydroxy-2-methylchromone formation using C1/C6 Claisen-type cyclization, but this mutant produced the other isomer 2,7-dihydroxy-5-methylchromone using C4/C9 aldol-type cyclization (Figure V.30) (Abe et al., 2005a).

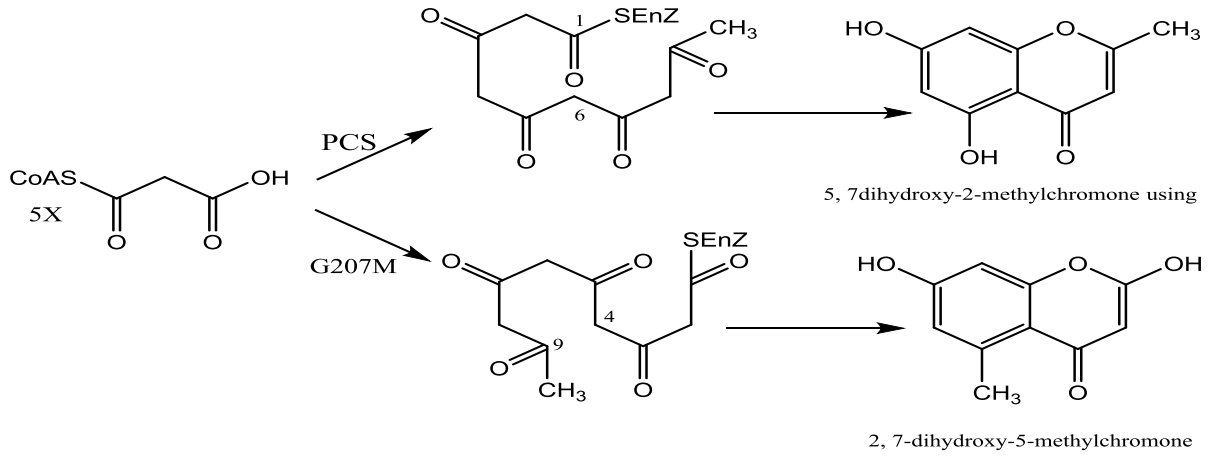


Fig V.30 The different aldol cyclization reactions of PCS and the G207M mutant of OKS.

1.3.7. The crystal structure of *Malus domestica* BIS3 complexed with benzoyl CoA

One of the collected datasets after X-ray measurement of the soaked crystals of MdBIS3 with benzoyl-CoA showed the presence of an electron density close to the catalytic Cys, which resembled the benzene ring. The total electron density in the active site cavity was modeled perfectly to benzoyl CoA. The residues of the substrate binding pocket aligned well with the substrate (Figure V.31).

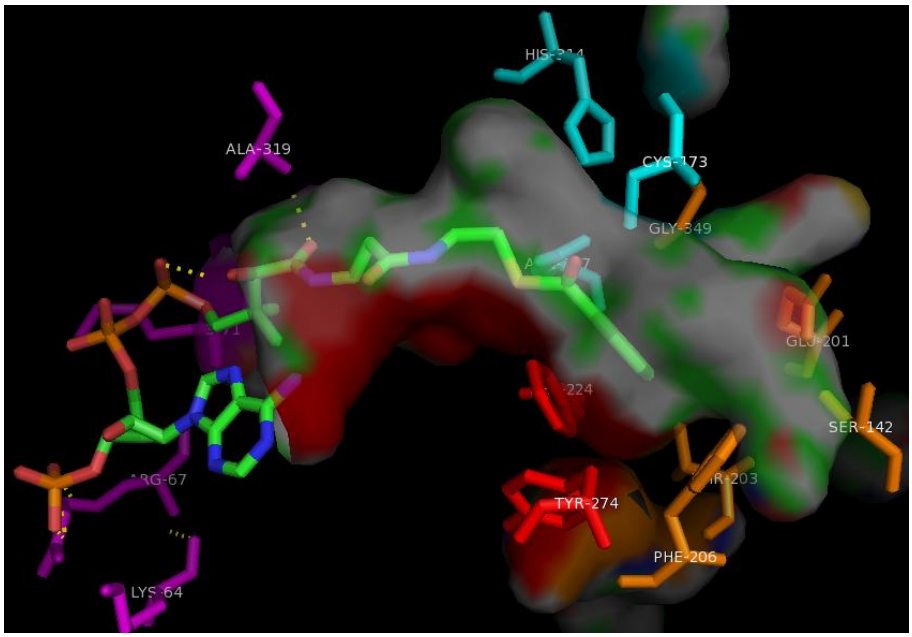


Fig V.31 The structure of MdBIS3 in complex with benzoyl-CoA (green). The residues of the catalytic triad (cyan), substrate binding pocket (orange), gatekeeper (red) and the CoA-binding tunnel (violet) are highlighted.

In comparison to the structure of the 18xCHS mutant complexed with resveratrol (PDB: 1U0W) (Austin et al., 2004a), we saw different orientation of the residues around the cavity. In addition, the hydoxyl group of the coumaroyl ring does not have enough place to be locked in the smaller cavity of MdBIS3 (Figure V.32).

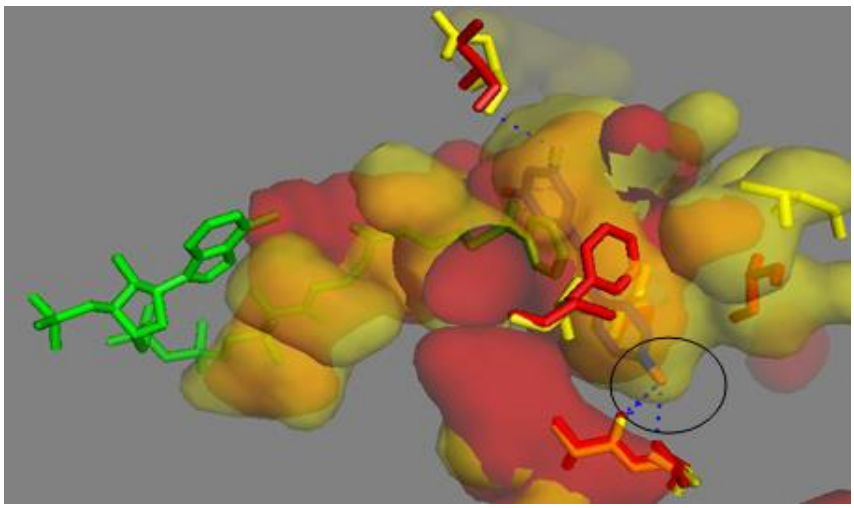


Fig V.32 The active site pockets of 18xCHS (yellow) complexed with resveratrol (blue) and MdBIS3 (red) complexed with benzoyl-CoA (green). The residues involved in polar interaction (blue dotted lines) with resveratrol are shown as sticks, like other major different residues of the substrate binding pockets.

Based on modeling benzoyl-CoA into the HaBPS pocket and comparison to the structure of MsCHS2 complexed with naringenin (PDB: 1CGK), we also see that the cavity of HaBPS is smaller and cannot accommodate the cyclized product of MsCHS2, especially at the region of its hydroxyl groups (Figure V.33).

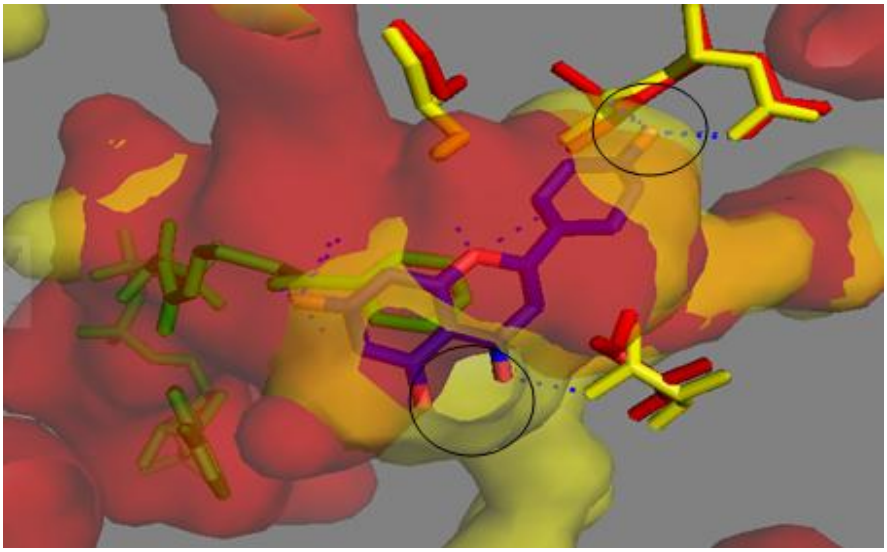


Fig V.33 The active site pockets of MsCHS2 (yellow) complexed with naringenine (blue) and HaBPS (red) modeled with benzoyl-CoA (green). The residues involved in polar interaction (blue dotted lines) with naringenine are shown as sticks, like the other major different residues of the substrate binding pockets.

Ultimately, structural differences among PKSs that alter the volume and the shape of the initiation/elongation cavity allow for discrimination between starter molecules and the number of elongation steps to direct the nature and length of the polyketide products (Jez et al., 2000a).

The cavities of HaBPS and the MdBIS3 are really like a reflection picture of their final products. Their shapes fit perfectly to the produced compounds, 3,5-dihdroxybiphenyl by MdBIS3 and 2,4,6-trihydroxybenzophenon by HaBPS (Figure V.34a, b).

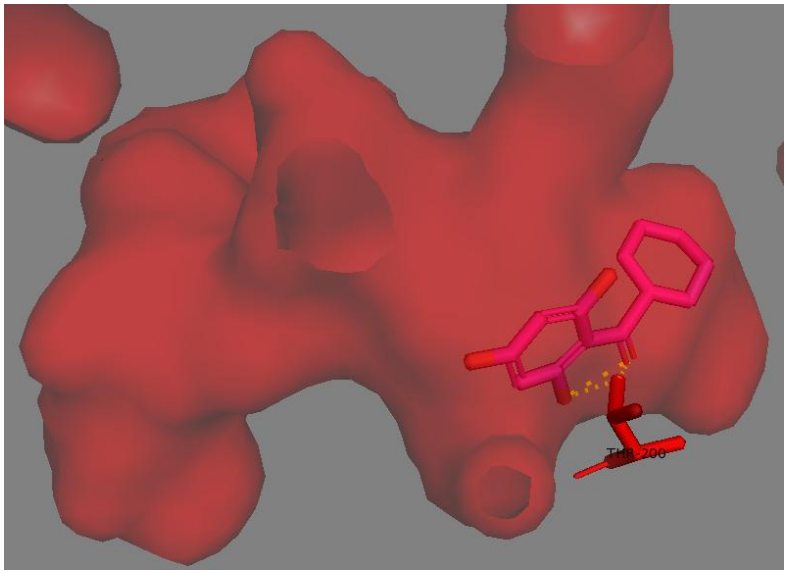


Fig V.34a The active site pocket of HaBPS modeled with 2,4,6-trihydroxybenzophenone (pink). The main different residue in the substrate binding pocket (Thr200) is marked as red sticks.

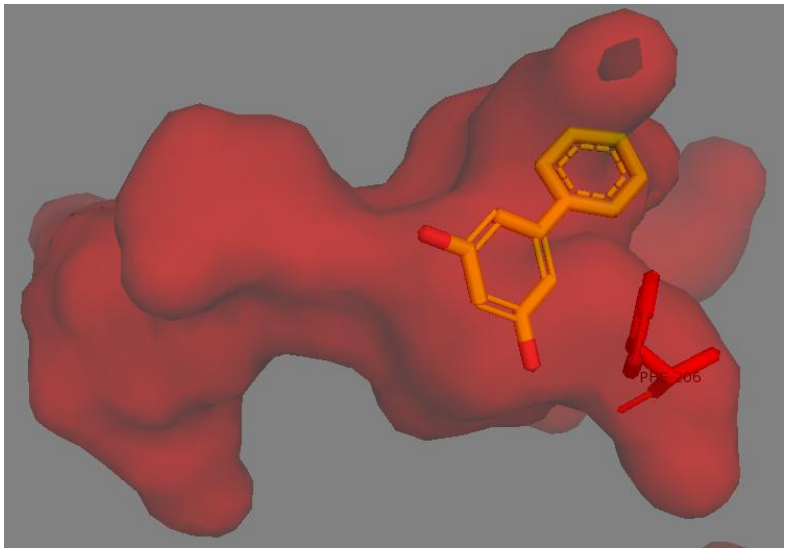


Fig V.34b The active site pocket of MdBIS3 modeled with 3,5-dihydroxybiphenyl (yellow). The main different residue in the substrate binding pocket (Phe192) is marked as red sticks.

1.3.8.Mutation and conversion of BPS into BIS

Another approach was performed to understand the reaction mechanisms of both enzymes. By transformation of BPS into BIS, we can exactly determine the important residues responsible for aldol or Claisen type cyclization. By comparing the crystal structures of HaBPS and MdBIS3 and analyzing the residues surrounding the active site pockets, we could determine many amino acids which showed different side chains and orientations between the two enzymes (Table 8). The 14 or the 13 mutations that we made were massive and they resulted in complete loss of the activity. Although the 14x or the 13x mutants enfolded the residues of the initiation/elongation cavities of both enzymes, there was neither BPS nor BIS activity. In each mutant there is one missing site from the elongation pocket as shown in the following table and it may be worth in the future to include these extra single sites into the constructed mutants (S219A in 14x mutant) and (M199V in 13x mutant), which may retain either BPS or BIS activity.

Table 8. Survey of mutations made to convert BPS into BIS

Enzyme	HaBPS	MdBIS3	14x mutant	13x mutant
Initiation pocket	Met217	Leu209	M217L	M217L
	Ile258	Val250	I258V	I258V
	Ala260	Ala252	F262I	F262I
	Tyr269	Tyr261	A257G	K272S
			F279V	F279V
Elongation pocket	Thr135	Ala127	T135A	T135A
	Ser136	Ser128	F138V	A133C
	Thr197	Thr189	N164E	N164E
	Met 199	Val91	M199V	Q165A
	Thr 200	Phe192	T200F	T200F
	Ser219	Ala211	Q165A	S219A
	Met267	Phe259	M267F	M267F
	Gly342	Gly334	F168Y	
			L377I	L377I

The previous conversion of CHS into STS also included 18 mutation sites (18x CHS) but the mutation process was gradual and stepwise (Austin et al., 2004a). Each time, there was an extra addition of another mutation site to the previously added one. Each single mutant was tested for its activity to perform exactly what was the function of the residue that was replaced. This strategy helped for the accurate understanding of the variation in the reaction mechanisms of CHS and STS and provided extra information to understand more about the variety in substrates acceptance and the different products formed by plant type III PKSs. The authors constructed about 81 four digit number mutants, in which each number indicates the partial (1) or the no (0) or the complete (2) mutation of one of the four important different areas between the two enzymes. For example the 2222 mutant is for the 18xCHS, 0002 for the mutant which included full mutation in area 4 but showed only CHS activity, which in turn indicated the unimportance of this area in contrast to area 2 in which 1210 mutant (8x CHS) showed STS like activity (Figure V.35) (Austin et al., 2004a). Looking in the structures of HaBPS and MdBIS3, this area 2 did not show notable displacement although it demonstrates different residues among the two enzymes.

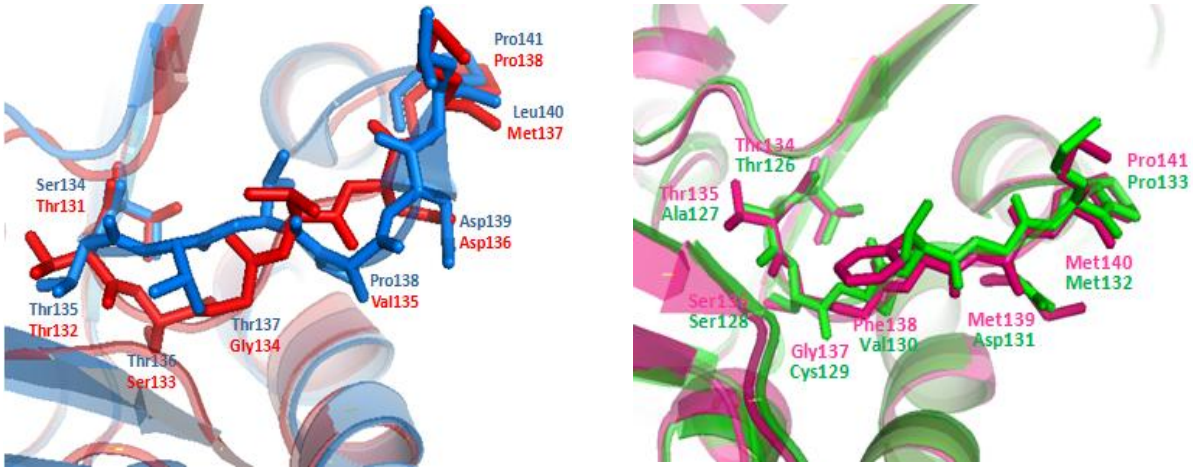


Fig V.35 The differences in area 2 loop among the four enzymes MsCHS (red), PsSTS (blue), HaBPS (pink) and MdBIS3 (green).

The previous difficult mutation strategy of MsCHS2 and PsSTS was performed in two enzymes, which produce trace amounts of the product of the other. This was reported in *Pueraria lobata* CHS that formed 2.7-4.2% resveratrol and AhSTS that formed 1.4-2.3% naringenine (Yamaguchi et al., 1999). In contrast, our enzymes BPS and BIS produce totally different products. For this reason, such a mutation strategy is worth being applied in the future. The conversion of STS into CHS is not reported either, therefore the mutation in the direction of BIS into BPS may be be feasible. However, the two enzymes share only 54% amino acids sequence similarity (Liu et al., 2007), which is very low in comparison to CHS and STS which have 75%-90% similarity (Austin et al., 2004a).

1.3.9. The crystal structures of HaBPS, HsBPS and the different surface residues

The structure of HsBPS was solved at 1.7 Å, displaying some changes in the geometry of the active sites pocket compared to that of HaBPS. The amino acids sequence alignment shows only six different amino acids (Asp54, Ala146, Gly230, Ala235, Leu359 and Ser360). Interestingly, they are far from the active site, but somehow they resulted in different biochemical and functional properties between the two enzymes (Figure V.36).

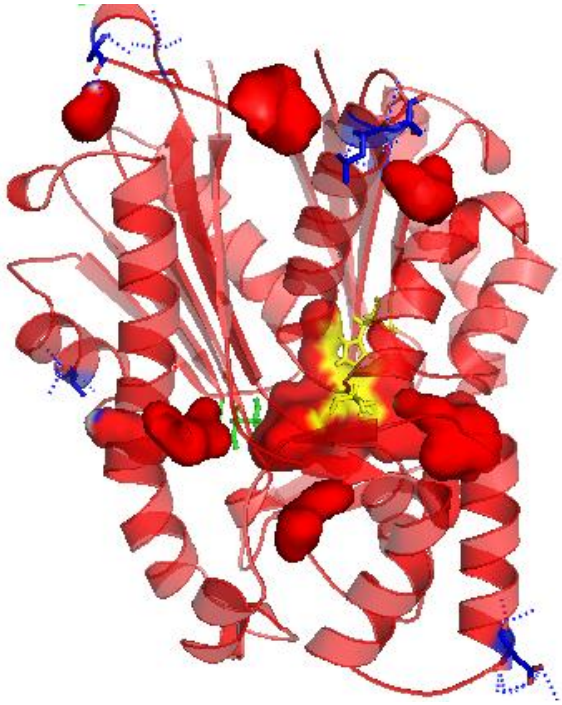


Fig V.36 The six different surface amino acids of HsBPS compared to HaBPS. The residues are shown as blue sticks together with Thr135 (green) in the center. Their polar interactions with other residues are shown as blue dotted lines.

The crystal structures of both enzymes demonstrate different shapes of the loops. These loop do not only surround the cyclization pockets but are also on the surface especially at the areas where there are different amino acids. The residues forming these loops take different orientations especially in their side chains. The residues of these loops in HaBPS are E83NQ (L1), G253SDGAITA (L2), Y269FLKEDVI (L3), P308GGRGI (L4), V359NG (L5), the famous G376LGPGL (L6) and I393LQ (L7) (Figure V.37). Most of the loops include or are close to an important residue (colored with red) of the active sites pocket. A260 corresponds to G256 of MsCHS2 and Y269 corresponds to F265 ofMsCHS2. Both residues, as mentioned previously, are known for their effects on substrate specificity and chain elongation. In addition, the loop L4 is close to His307 of the catalytic triad. The loop L5 contains two surface residues different from HsBPS, V359 and N360.

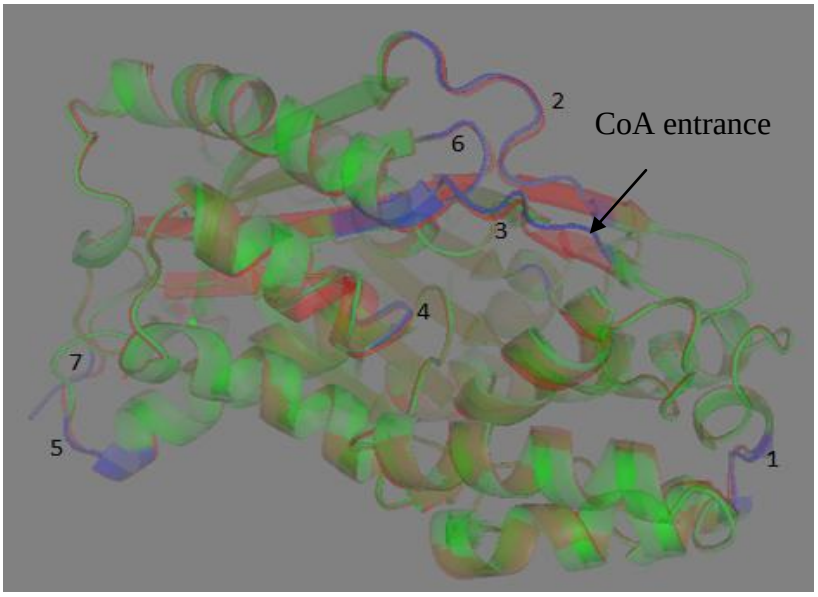


Fig V.37 The differently diplaced loops of HaBPS (green) and HsBPS (red). The main different areas are highlighted in blue. The numbers indicate the previously mentioned loop numbers.

These loops surround mainly the entrance of the active sites pocket, which should correlate to the differences in the substrate specificities, as HsBPS accepts salicoyl-CoA and HaBPS accepts N-methylanthraniloyl-CoA. Analyzing these loops in detail, we see a specific area of some similar residues in loop L4 which have different orientations between the two enzymes. These are Gly309, G310 and Ile313. The amine and carbonyl groups of both Gly residues are oriented in the same directions but highly displaced from each other. The side chain of Ile313 is directed toward the active sites pocket of HaBPS and in HsBPS it moves away from the cavity (Figure V.38).

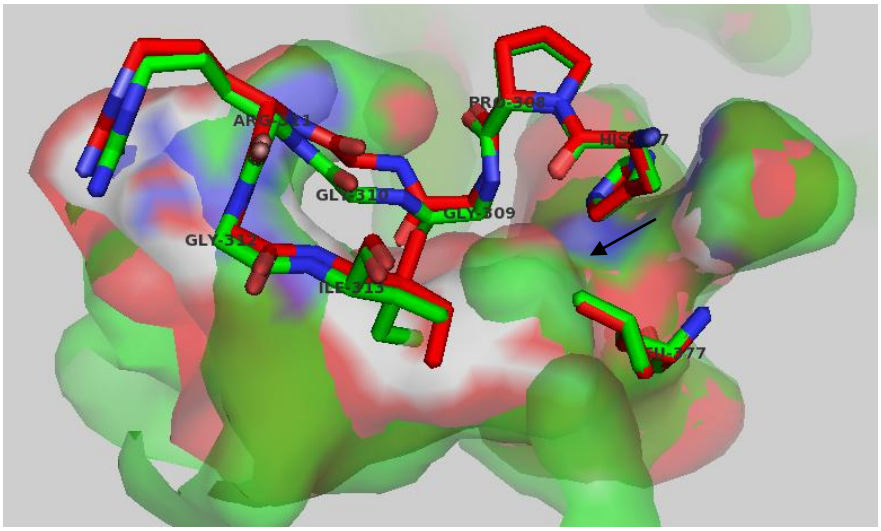


Fig V.38 The specifically displaced loop of HaBPS (green) and HsBPS (red). The loop surrounds the CoA tunnel. The side chain of Ile313 in HaBPS directs toward the active site pocket, in contrast to Ile313 of HsBPS. The slightly expanded pocket (marked with arrow) is located in the middle between Ile313 and Leu377.

The previous three residues also interact with Gly312, which shows also slight displacement. This residue is one of the important residues of the CoA binding tunnel. There is a slightly expanded pocket in the surface between the three residues (Gly309, Gly310 and Ile313) and Leu377. This slightly expanded pocket resulted from the displacement of these residues between the two enzymes and it is bigger in HaBPS than that in HsBPS. These variations are at the area of the CoA binding tunnel and close to the substrate binding pocket and thus should correlate to the differences in the substrate specificities of HaBPS and HsBPS. The absence of structures of enzyme-substrate or enzyme-product complexes hinders the interpretation of this area at the CoA tunnel. Until now, there are no other type III PKS structures that are complexed with salicoyl-CoA, but in the PDB databank we found one PKS-like enzyme (PqsD) from *Pseudomonas aeruginosa*, the structure of which is complexed with anthranilic acid. Anthraniloyl-CoA is the first of two substrates used by PqsD. CoA delivers anthranilate to the PqsD active sites Cys112. PqsD uses malonyl-CoA and malonyl-ACP as its second substrate in a reaction that forms 2,4-dihydroxyquinoline (DHQ). It shows the dimer structure as type III PKSs. The structure demonstrates a similar slightly expanded pocket that locks the aminophenyl ring as the one that we found in HaBPS and HsBPS (Figure V.39).

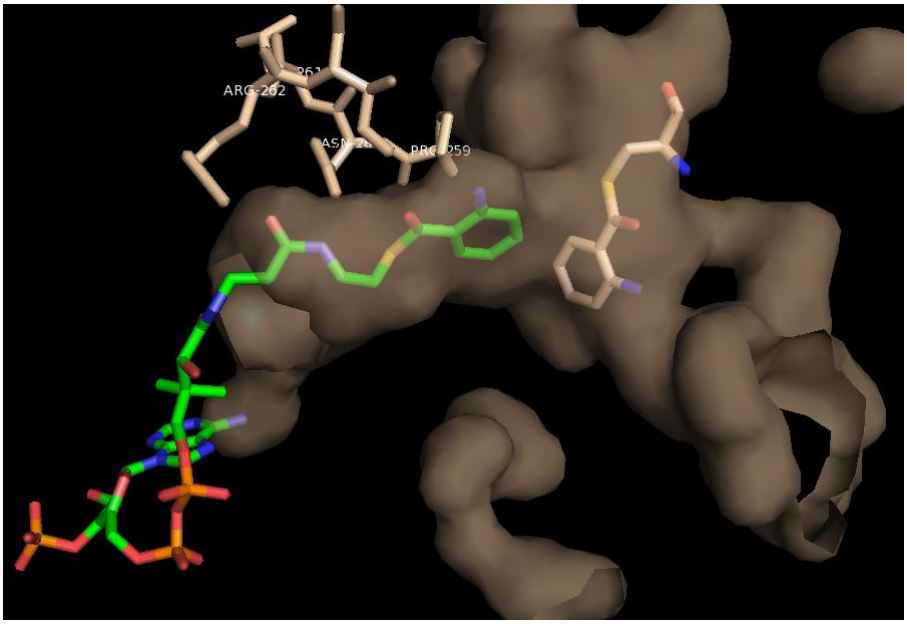


Fig V.39 The crystal structure of PqsD from *Pseudomonas aeruginosa* in complex with anthranilic acid (pink) and anthraniloyl-CoA (green). The residues of the loop above the CoA tunnel are shown as sticks.

Comparison of the PqsD-anthranilate complex structure with our structures of HaBPS and HsBPS, let us, assume that, the elongations direction of the growing polyketide chain using anthraniloyl or N-methylanthraniloyl-CoA is downward. The bottom of the active sites pocket is not large enough for several elongations using this bulk substrate, therefore only one molecule of malonyl-CoA is added to result in the corresponding quinolone. On the other hand, the elongation with benzoyl-CoA will move toward the normal big benzoyl-binding pocket in the back of the cavity, which will allow for the addition of three molecules of malonyl-CoA.

These directions of elongations are likely the same in case of MdBIS3 and salicoyl-CoA, as this substrate is also substituted at ortho-position and undergoes only one condensation with malonyl-CoA. As this elongation area is smaller in MdBIS3 than in HaBPS, it accepts salicoyl-CoA but not anthraniloyl- or N-methylanthraniloyl-CoA. Regarding HaBPS and HsBPS, this area is nearly the same but it is still not clear why salicoly-CoA is utilized by HsBPS and not by HaBPS. Those are still predictions depending on the bacterial PqsD-anthranilate complex structure because of the absence of complex structures for the plant type III PKSs, which use anthraniloyl- or N-methylanthranuloyl-CoA as their natural substrates, like CmACS and CmQNS. Since there are high similarities between HaBPS and CmACS as well as MdBIS3 and CmQNS, as mentioned previously, there are great chances to transform benzoic acid-specific type III PKSs into ACS or QNS using mutagenesis (Figure V.40).

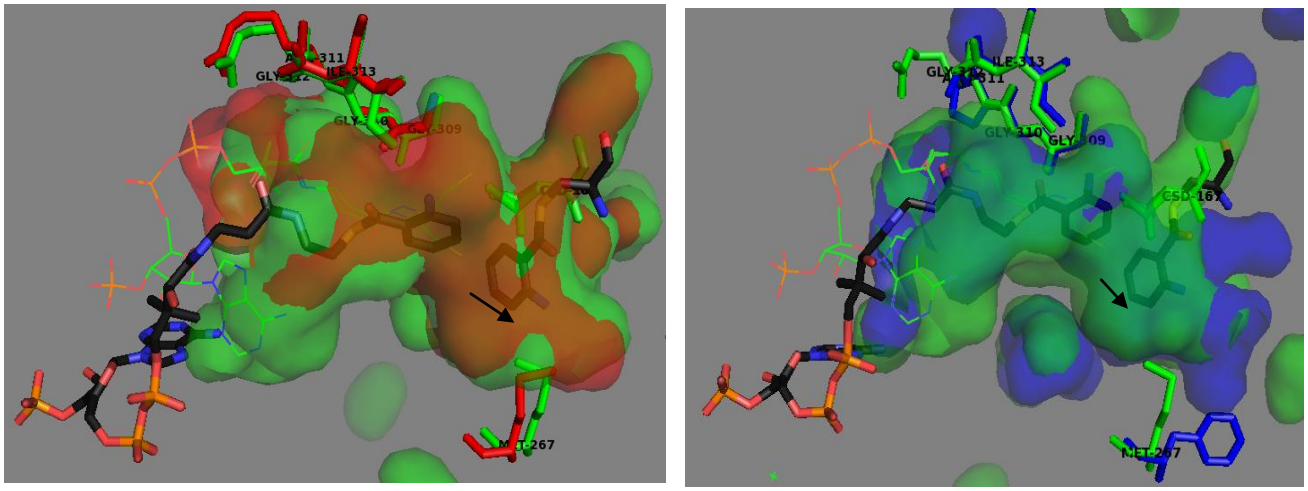


Fig V.40 The structures of HaBPS (green), HsBPS (red) and MdBIS3 (blue) modeled with anthranilic acid. Anthraniloyl-CoA and anthranilic acid are shown as black sticks. The CoA molecule of the HaBPS-CoA structure is shown as lines. The arrows indicate the suggested pocket, which locks the resulting quinolone. This pocket is surrounded by Met267, which shows clear displacement in HaBPS and HsBPS and is replaced with Phe259 in MdBIS3. The different residues of the specific loop above the CoA tunnel are shown as sticks. The specific orientation of Arg311, which corresponds to Pro303 in MdBIS3, is shown in both BPSs.

Met267 shows different orientation in HaBPS and HsBPS; the location of this residue is close to the pocket that probably locks the produced quinolone. Comparing the top of the CoA binding tunnel of HaBPS and HsBPS to that of MdBIS3, there is one interesting residue, Arg311 of BPSs, which corresponds to Pro303 of MdBIS3. The pyrrolidine ring of Pro303 is directed toward the cavity and its existence may sterically hinder the amine or N-methylamine groups of anthraniloyl- or N-methylanthranuloyl-CoA, respectively. In contrast, Arg311 of both BPSs moves away from the cavity to allow the big side chains of the previously mentioned substrates to enter easily. Mutagenesis of this residue will indicate if it has effect on substrate acceptance, especially when mentioning that CmQNS has Arg307 at this site but CmACS has Pro307 instead.

Mutational experiments were conducted to convert HaBPS to HsBPS by substitution of the six surface amino acids (Glu54, Val146, Cys230, Ser235, Val359 and Asn360) to the corresponding ones of HsBPS (Asp54, Ala146, Gly230, Ala235, Leu359 and Ser360). HsBPS demonstrated high *Km* values for both the starter and extender substrates in comparison to other *Hypericum* BPSs. The previously produced single mutants of HaBPS showed also 3 to 4.9 and 2 to 4-fold increases in the malonyl-CoA and benzoyl-CoA *Km* values, respectively, compared to the wild-type enzyme. All the previous single mutants were active but only HaS235A did not show any over-expression (Abdelaziz, 2014). The result of no protein expression after surface amino acids mutation is actually expected. For example, the surface residues Arg199 and Arg350 in CHS from *Pueraria lobata* have found to maintain the structure integrity, as the side chain of Arg199 forms hydrogen bonds with the carbonyl oxygens of both Val261 and Leu263. In addition, the two nitrogen atoms of Arg350 are close to the oxygen atoms of highly conserved Glu116. Moreover the produced mutants of these two residues resulted in decreased yields of protein amount, which means that these Args also affect the protein folding (Fukuma et al., 2007).

In HaBPS, the loop Thr135 to Tyr146 is located at the dimer interface but does not show differences between HaBPS, HsBPS and even HaT135L. The loop Gly230 to Ala235 is located on the surface and shows slight displacement among HaBPS, HsBPS and HaT135L. This region also differs in the sequences of *H. calycinum* and *H. perforatum* BPSs. Close to this area is the N-terminal helix 1 – 20, which differs in *H. perforatum* ssp *angustifolium*, and it also shows displacement HaBPS, HsBPS and HaT135L. The helix Val346 to Asn360 is running into the active site but does not show any differences between HaBPS, HsBPS and even HaT135L. Glu54 is the closest residue to the CoA tunnel and its orientation is similar in HaBPS, HsBPS and HaT135L. Those loops which included the six different amino acids between HaBPS and HsBPS did not demonstrate big changes as the displaced loops that we identified earlier by the crystal structures of both enzymes (Figure V.41).

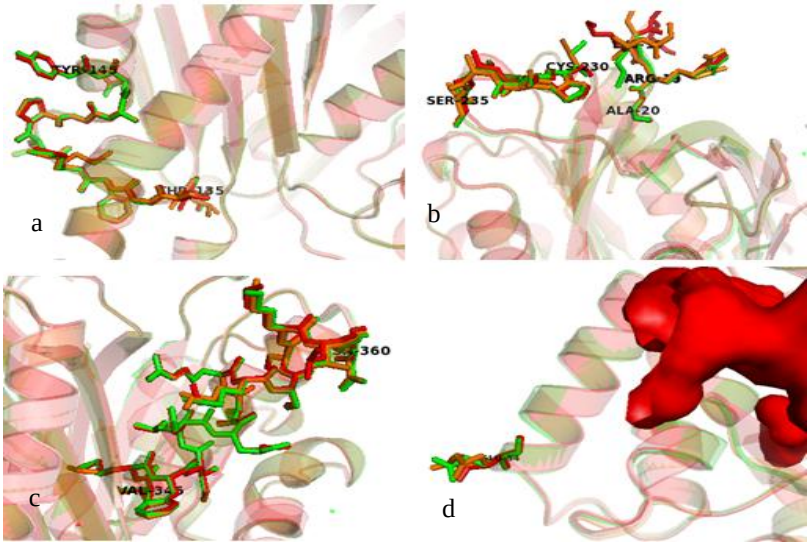


Fig V.41 The loops that include the six surface amino acids, which differ between HaBPS (green) and HsBPS (red). a) loop T135-Y146, b) loop G230-A235 with N-terminal helix 1-20, c) helix V346-N360, d) E54.

In contrast to HaBPS, two specific mutants, HsBPST135I and HsBPST135K, resulted in the transformation of HsBPS into phenylpyrone synthase and hydroxycoumarin synthase, respectively. Interestingly, the double mutant HaBPSS235A/T135I was active and resulted in phenylpyrone formation, although the single site mutant HaBPST135I was inactive in comparison to the active HsBPST135I (Abdelaziz, 2014). Even by analyzing the mutation area of both enzymes, there was no clear difference in the orientation of the residues surrounding Thr135. These results strongly suggest the high impact of the six surface amino acids on other important residues surrounding the active site cavity such as Thr135. The crystal structure of HsBPS demonstrates a large network between these six surface amino acids and the Thr135 beside the involvement of other residues which maintain the integrity and the geometry of the active site pocket, such as the G376LGPG loop corresponding to the G372FGPG loop of MsCHS2 (Figure V.42).

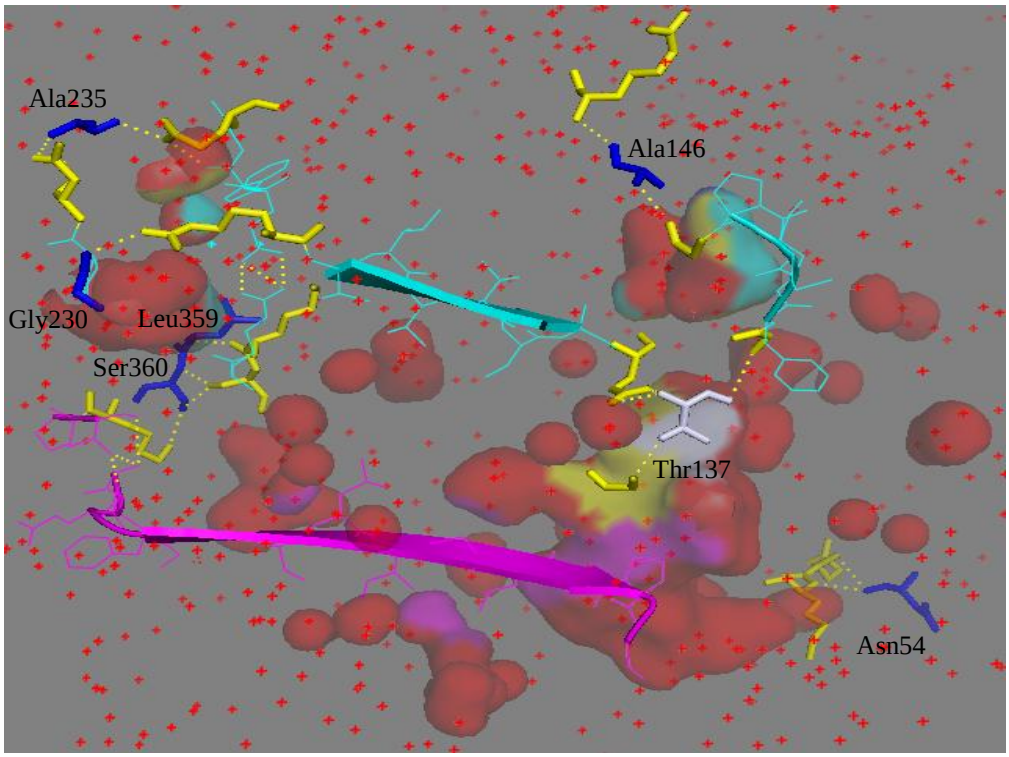


Fig V.42 The six surface residues of HsBPS (blue) involved in hydrogen bonding with other residues (yellow). These residues together maintain the shape and the integrity of the active site pocket (red) by taking part in important loops (cyan), such as the G376LGPG loop (purple). This network also involves interaction with Thr135 (white).

Mutational experiments proved the importance of the flexibility of the G372FGPG loop in polyketide formation, as it takes part in the cyclization pocket of type III PKSs (Suh et al., 2000). There is also a relationship between this loop and some surface amino acids. For example, when a mutation was conducted on CHS from *Pueraria lobata* to study the effect of some surface arginines on the enzyme acivity, Arg68, Arg172 and Arg328 turned out to help in the correct positioning of the catalytic Cys-His-Asn triad and the G$_{372}$FGPG loop.

The Arg172 mutant showed dramatic increase in the CTAL production, as this residue is located close to the catalytic nucleophile Cys164 and the G_{372}FGPG loop. It helps in the correct orientation of the loop toward the catalytic Cys164, which affects the cyclization reaction of the enzyme (Fukuma et al., 2007).

Another important residue was found to have different orientation in HaBPS and HsBPS, which is Leu271 corresponding to Leu267 of MsCHS2. This residue shows different orientations in HaBPS, HsBPS and HaT135L. In addition, the loop close to it and just above the CoA binding tunnel shows clear displacement in HaBPS, HsBPS and also HaBPST135L. In CmACS, the conformational change of Leu267 together with Phe215 was the reason of widening its active site entrance. The broad active site entrance affects the starter substrate preference and the product specificity of CmACS (Figure V.43) (Mori et al., 2013).

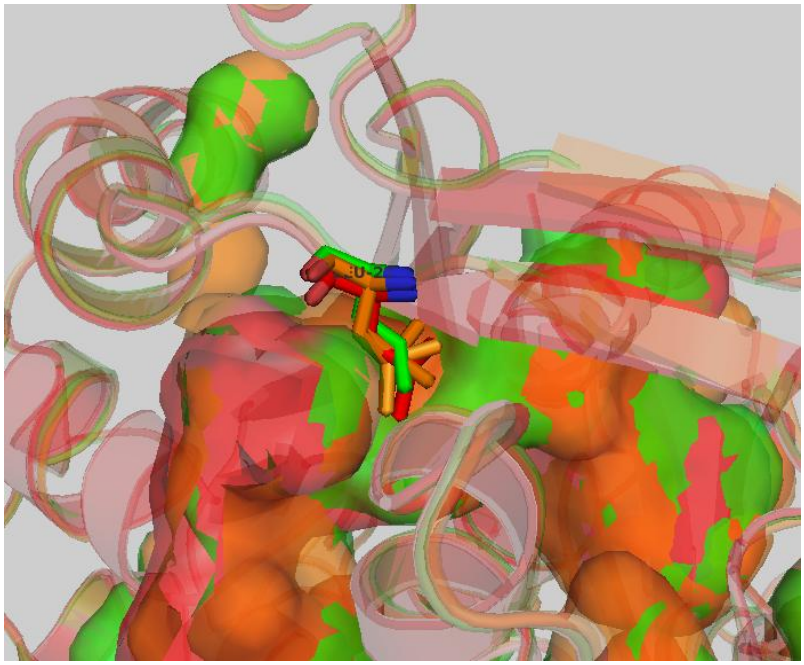


Fig V.43 The different orientations of Leu271 in HaBPS (green), HsBPS (red) and HaBPST135L. The close loop to Leu271 and just above the CoA tunnel shows displacement in all three enzymes.

All these previously mentioned changes were due to the six different surface residues in HaBPS and HsBPS. In the future, surface residues should be the target for mutational experiments, as they showed in many cases an important effect on the activity of plant type III PKSs.

VI. Summary

- Type III PKSs share a common αβαβα-fold and a conserved catalytic triad, Cys-His-Asn, in their three-dimensional structures. The identity and position of each amino acid residue of the active site cavity account for the differences observed between type III PKS activities.
- Benzoic acid-specific type III PKSs constitute a very interesting group because of their close similarities to type II PKSs, which synthesize a serious of interesting medicinal aromatic products, such as antibiotics. However, in contrast to type II PKSs, they are smaller and simpler in their structures. They utilize benzoyl-CoA as the preferred starter substrate.
- Benzophenone synthase (BPS) is responsible for the formation of the C_{13} skeleton of many phenolic natural products with interesting pharmacological activities. Biphenyl synthase (BIS) synthesizes biphenyls and, derived thereof, dibenzofurans which have antimicrobial activities. BPS and BIS catalyze iterative condensation of benzoyl-CoA with three molecules of malonyl-CoA to form a linear tetraketide. This intermediate undergoes C6 ⟶ C1 Claisen condensation and aromatization by BPS to form trihydroxybenzophenones. In case of BIS, it undergoes C2 ⟶ C7 aldol condensation and aromatization to form dihydroxybiphenyls.
- Crystal structures of many plant and some microbial type III PKSs have recently been solved. Although, many different scenarios were hypothesized as an explanation for the two different reaction mechanisms of these enzymes, it remains open if either electronic or steric factors are the major reasons behind the difference. Despite that, it is clear that minor structural changes can direct the natural mechanism of one enzyme toward that of the other one, which leads to formation of novel unnatural products.
- Cleavage of the His_6-tag was improved at different temperatures and time intervals. The best cutting conditions using TEV protease were at 37 °C for 4 h. However, gel filtration chromatography could not separate the cutted and the non-cutted enzyme completely. Therefore, the crystallization process was carried out using the full His_6-tagged protein.
- Protein expression and purification optimizations were established using high culture volumes for *H. androsaemum* BPS (HaBPS), its mutant HaBPST135L, *H. sampsonii* BPS (HsBPS), and *M. domestica* BIS3 (MdBIS3). Crystallization of all these enzymes was carried out using the vapor diffusion method in close cooperation with the Department of Structure and Function of Proteins**,** Helmholtz Centre for Infection Research, Braunschweig, which provided the buffer screens as well as the expertise and equipment needed for protein crystallization.
- The crystal structures of HaBPS and its mutant HaBPST135L were solved at 2.1 Å and 1.4 Å, respectively. The orientation of Leu135 proved the formation of a new pocket, which was explained previously by structure modelling of the mutant. The active site pocket in HaBPST135L is smaller than in HaBPS, which explains the production of a triketide chain rather than a tetraketide chain by the wild-type enzyme.

- The crystal structure of MdBIS3 was solved at 1.9 Å. It identified a new active site pocket, which resembled that of HaBPST135L and seems to be a characteristic pocket of benzoic acid-specific type III PKSs. The side chains of Ala127 in MdBIS3 and Thr135 in HaBPS line this novel pocket.
- Claisen versus aldol cyclization differences of type III PKSs are mostly correlated to the presence of a water molecule in their crystal structures. Actually, a specific water molecule was found in the new pocket of HaBPS and MdBIS3. It is exactly present at the position of the water molecule that was responsible for the aldol-switch mechanism of *P. sylvestris* STS. The orientations of Thr135 and Gly166 of HaBPS influence the geometry of the neighbouring residues and eventually the hydrogen bond network around the new water molecule.
- Oxidized catalytic Cys was observed in HaBPS and MdBIS3, which is a well-known phenomenon of type III PKSs. Therefore, soaking and co-crystallization were commonly carried out using the active site mutants HaBPSC167A and MdBIS3C159A. Benzoyl-CoA was modeled in one of the datasets of soaked MdBIS3 crystals.
- The crystal structure of HsBPS was solved at 1.7 Å. In comparison to HaBPS, it has many displaced loops, which not only surround the active site pocket but also enclose surface areas with amino acids differing between the two enzymes. The side chains of the residues forming these loops have different orientations. Structures of enzyme-substrate and/or enzyme-product complexes are needed for detailed interpretation of both structures. Crystallization of active site mutants of HsBPS, HsBPST135K and HsT135I, is important to study structural differences in comparison to non-active mutants of HaBPS.
- Transformation of BPS into BIS is able to identify the main residues that are responsiple for the different condensation and cyclization mechanisms. However, a 14-fold mutant of BPS exhibited neither BIS nor BPS activity. To achieve a transformation toward BPS in the future, a mutagenesis strategy should be carried out, which starts with mutations of BIS or even the BPST135L mutant that has a similar new pocket like BIS. In addition, interesting newly cloned BPSs and BISs may be available in the future for crystallization, using the conditions optimized for HaBPS and MdBIS3, thereby increasing the chance of solving the mystery of the aldol/Claisen cyclization mechanism variation.
- The active site pockets of HaBPS and MdBIS3 demonstrate high similarities in their geometry to two important alkaloid-synthesizing plant type III PKSs, namely *C. microcarpa* ACS and *C. microcarpa* QNS. In a phylogenetic tree, they are also close to CmQNS. This will provide the great chance of transforming benzoic acid-specific type III PKSs into ACS and QNS using mutagenesis.

VII. References

Abdelaziz S, 2014. *Site-directed mutagenesis of Hypericum benzophenone synthases* TU Braunschweig.

Abe I, 2008. Engineering of plant polyketide biosynthesis. *Chem Pharm Bull (Tokyo)* **56**, 1505-14.

Abe I, Abe T, Wanibuchi K, Noguchi H, 2006a. Enzymatic formation of quinolone alkaloids by a plant type III polyketide synthase. *Org Lett* **8**, 6063-5.

Abe I, Morita H, 2010. Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. *Nat Prod Rep* **27**, 809-38.

Abe I, Morita H, Oguro S, *et al.*, 2007a. Structure-based engineering of a plant type III polyketide synthase: formation of an unnatural nonaketide naphthopyrone. *J Am Chem Soc* **129**, 5976-80.

Abe I, Oguro S, Utsumi Y, Sano Y, Noguchi H, 2005a. Engineered biosynthesis of plant polyketides: chain length control in an octaketide-producing plant type III polyketide synthase. *J Am Chem Soc* **127**, 12709-16.

Abe I, Sano Y, Takahashi Y, Noguchi H, 2003. Site-directed mutagenesis of benzalacetone synthase. The role of the Phe215 in plant type III polyketide synthases. *J Biol Chem* **278**, 25218-26.

Abe I, Takahashi Y, Morita H, Noguchi H, 2001. Benzalacetone synthase. A novel polyketide synthase that plays a crucial role in the biosynthesis of phenylbutanones in Rheum palmatum. *Eur J Biochem* **268**, 3354-9.

Abe I, Utsumi Y, Oguro S, Morita H, Sano Y, Noguchi H, 2005b. A plant type III polyketide synthase that produces pentaketide chromone. *J Am Chem Soc* **127**, 1362-3.

Abe I, Watanabe T, Lou W, Noguchi H, 2006b. Active site residues governing substrate selectivity and polyketide chain length in aloesone synthase. *FEBS J* **273**, 208-18.

Abe T, Morita H, Noma H, Kohno T, Noguchi H, Abe I, 2007b. Structure function analysis of benzalacetone synthase from Rheum palmatum. *Bioorg Med Chem Lett* **17**, 3161-6.

Austin MB, Bowman ME, Ferrer JL, Schroder J, Noel JP, 2004a. An aldol switch discovered in stilbene synthases mediates cyclization specificity of type III polyketide synthases. *Chem Biol* **11**, 1179-94.

Austin MB, Izumikawa M, Bowman ME, *et al.*, 2004b. Crystal structure of a bacterial type III polyketide synthase and enzymatic control of reactive polyketide intermediates. *J Biol Chem* **279**, 45162-74.

Austin MB, Noel JP, 2003. The chalcone synthase superfamily of type III polyketide synthases. *Nat Prod Rep* **20**, 79-110.

Baumert A, Maier W, Groger D, Deutzmann R, 1994. Purification and properties of acridone synthase from cell suspension cultures of Ruta graveolens L. *Z Naturforsch C* **49**, 26-32.

Beerhues L, 1996. Benzophenone synthase from cultured cells of Centaurium erythraea. *FEBS Lett* **383**, 264-6.

Bergfors T, 2003. Seeds to crystals. *J Struct Biol* **142**, 66-76.

Birnboim HC, Doly J, 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**, 1513-23.

Birt DF, Widrlechner MP, Hammer KD, *et al*., 2009. Hypericum in infection: Identification of anti-viral and anti-inflammatory constituents. *Pharm Biol* **47**, 774-82.

Bisht K, Wagner KH, Bulmer AC, 2010. Curcumin, resveratrol and flavonoids as anti-inflammatory, cyto- and DNA-protective dietary compounds. *Toxicology* **278**, 88-100.

Bradford MM, 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248-54.

Bruni R, Pellati F, Bellardi MG, *et al*., 2005. Herbal drug quality and phytochemical composition of Hypericum perforatum L. affected by ash yellows phytoplasma infection. *J Agric Food Chem* **53**, 964-8.

Carson M, Johnson DH, Mcdonald H, Brouillette C, Delucas LJ, 2007. His-tag impact on structure. *Acta Crystallogr D Biol Crystallogr* **63**, 295-301.

Chemler JA, Buchholz TJ, Geders TW, *et al*., 2012. Biochemical and structural characterization of germicidin synthase: analysis of a type III polyketide synthase that employs acyl-ACP as a starter unit donor. *J Am Chem Soc* **134**, 7359-66.

Chen JJ, Chen HJ, Lin YL, 2014. Novel polyprenylated phloroglucinols from Hypericum sampsonii. *Molecules* **19**, 19836-44.

Chizzali C, Gaid MM, Belkheir AK, *et al*., 2012. Differential expression of biphenyl synthase gene family members in fire-blight-infected apple 'Holsteiner Cox'. *Plant Physiol* **158**, 864-75.

Crockett SL, Robson NK, 2011. Taxonomy and Chemotaxonomy of the Genus Hypericum. *Med Aromat Plant Sci Biotechnol* **5**, 1-13.

Du L, Zhu T, Liu H, Fang Y, Zhu W, Gu Q, 2008. Cytotoxic polyketides from a marine-derived fungus Aspergillus glaucus. *J Nat Prod* **71**, 1837-42.

Duff MR, Jr., Grubbs J, Howell EE, 2011. Isothermal titration calorimetry for measuring macromolecule-ligand affinity. *J Vis Exp*.

Ferrer JL, Jez JM, Bowman ME, Dixon RA, Noel JP, 1999. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat Struct Biol* **6**, 775-84.

Flores-Sanchez IJ, Verpoorte R, 2009. Plant polyketide synthases: a fascinating group of enzymes. *Plant Physiol Biochem* **47**, 167-74.

Fukuma K, Neuls ED, Ryberg JM, Suh DY, Sankawa U, 2007. Mutational analysis of conserved outer sphere arginine residues of chalcone synthase. *J Biochem* **142**, 731-9.

Funa N, Awakawa T, Horinouchi S, 2007. Pentaketide resorcylic acid synthesis by type III polyketide synthase from Neurospora crassa. *J Biol Chem* **282**, 14476-81.

Funa N, Ohnishi Y, Fujii I, Shibuya M, Ebizuka Y, Horinouchi S, 1999. A new pathway for polyketide synthesis in microorganisms. *Nature* **400**, 897-9.

Ghisalberti EL, 1996. Bioactive acylphloroglucinol derivatives from Eucalyptus species. *Phytochemistry* **41**, 7-22.

Gokulan K, O'leary SE, Russell WK, *et al.*, 2013. Crystal structure of Mycobacterium tuberculosis polyketide synthase 11 (PKS11) reveals intermediates in the synthesis of methyl-branched alkylpyrones. *J Biol Chem* **288**, 16484-94.

Gross F, Luniak N, Perlova O, *et al.*, 2006. Bacterial type III polyketide synthases: phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expression in pseudomonads. *Arch Microbiol* **185**, 28-38.

Hashimoto M, Nonaka T, Fujii I, 2014. Fungal type III polyketide synthases. *Nat Prod Rep* **31**, 1306-17.

Hopwood DA, 1997. Genetic Contributions to Understanding Polyketide Synthases. *Chem Rev* **97**, 2465-98.

Huang L, Wang H, Ye H, *et al.*, 2012. Differential expression of benzophenone synthase and chalcone synthase in Hypericum sampsonii. *Nat Prod Commun* **7**, 1615-8.

Jez JM, Austin MB, Ferrer J, Bowman ME, Schroder J, Noel JP, 2000a. Structural control of polyketide formation in plant-specific polyketide synthases. *Chem Biol* **7**, 919-30.

Jez JM, Bowman ME, Noel JP, 2001. Structure-guided programming of polyketide chain-length determination in chalcone synthase. *Biochemistry* **40**, 14829-38.

Jez JM, Ferrer JL, Bowman ME, Dixon RA, Noel JP, 2000b. Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. *Biochemistry* **39**, 890-902.

Junghanns KT, Kneusel RE, Baumert A, Maier W, Groger D, Matern U, 1995. Molecular cloning and heterologous expression of acridone synthase from elicited Ruta graveolens L. cell suspension cultures. *Plant Mol Biol* **27**, 681-92.

Katsuyama Y, Kita T, Funa N, Horinouchi S, 2009. Curcuminoid biosynthesis by two type III polyketide synthases in the herb Curcuma longa. *J Biol Chem* **284**, 11160-70.

Katsuyama Y, Matsuzawa M, Funa N, Horinouchi S, 2007. In vitro synthesis of curcuminoids by type III polyketide synthase from Oryza sativa. *J Biol Chem* **282**, 37702-9.

Katsuyama Y, Miyazono K, Tanokura M, Ohnishi Y, Horinouchi S, 2011. Structural and biochemical elucidation of mechanism for decarboxylative condensation of beta-keto acid by curcumin synthase. *J Biol Chem* **286**, 6659-68.

Khan N, Afaq F, Mukhtar H, 2008. Cancer chemoprevention through dietary antioxidants: progress and promise. *Antioxid Redox Signal* **10**, 475-510.

Klundt T, 2008. *Ortsgerichtete Mutagenese der Benzophenonsynthase von Hypericum androsaemum* Technische Universitaet Braunschweig.

Klundt T, Bocola M, Lutge M, Beuerle T, Liu B, Beerhues L, 2009. A single amino acid substitution converts benzophenone synthase into phenylpyrone synthase. *J Biol Chem* **284**, 30957-64.

Laemmli UK, 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-5.

Lanz T, Tropf S, Marner FJ, Schroder J, Schroder G, 1991. The role of cysteines in polyketide synthases. Site-directed mutagenesis of resveratrol and chalcone synthases, two key enzymes in different plant-specific pathways. *J Biol Chem* **266**, 9971-6.

Li J, Luo Y, Lee JK, Zhao H, 2011. Cloning and characterization of a type III polyketide synthase from Aspergillus niger. *Bioorg Med Chem Lett* **21**, 6085-9.

Liu B, Beuerle T, Klundt T, Beerhues L, 2004. Biphenyl synthase from yeast-extract-treated cell cultures of Sorbus aucuparia. *Planta* **218**, 492-6.

Liu B, Falkenstein-Paul H, Schmidt W, Beerhues L, 2003. Benzophenone synthase and chalcone synthase from Hypericum androsaemum cell cultures: cDNA cloning, functional expression, and site-directed mutagenesis of two polyketide synthases. *Plant J* **34**, 847-55.

Liu B, Raeth T, Beuerle T, Beerhues L, 2007. Biphenyl synthase, a novel type III polyketide synthase. *Planta* **225**, 1495-503.

Liu B, Raeth T, Beuerle T, Beerhues L, 2010. A novel 4-hydroxycoumarin biosynthetic pathway. *Plant Mol Biol* **72**, 17-25.

Lo MC, Aulabaugh A, Jin G, *et al.*, 2004. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal Biochem* **332**, 153-9.

Lukacin R, Schreiner S, Matern U, 2001. Transformation of acridone synthase to chalcone synthase. *FEBS Lett* **508**, 413-7.

Lütge M, 2012. *Untersuchungen zur Struktur und Funktion ausgewählter Polyketidsynthasen.*

Maier U-G, 1992. Gentechnische Methoden. Eine Sammlung von Arbeitsanleitungen für das molekularbiologische Labor. Hrsg. von Sabine Bertram und Hans Günter Gassen. Gustav Fischer Verlag, Stuttgart 1991. 326 S., 74 Abb., 20 Tabellen. DM 58,—. ISBN 3-437-2047-2. *Biologie in unserer Zeit* **22**, 59-.

Mcpherson A, Gavira JA, 2014. Introduction to protein crystallization. *Acta Crystallogr F Struct Biol Commun* **70**, 2-20.

Meslet-Cladiere L, Delage L, Leroux CJ, *et al.*, 2013. Structure/function analysis of a type iii polyketide synthase in the brown alga Ectocarpus siliculosus reveals a biochemical pathway in phlorotannin monomer biosynthesis. *Plant Cell* **25**, 3089-103.

Moore BS, Hopke JN, 2001. Discovery of a new bacterial polyketide biosynthetic pathway. *Chembiochem* **2**, 35-8.

Mori T, Shimokawa Y, Matsui T, *et al.*, 2013. Cloning and structure-function analyses of quinolone- and acridone-producing novel type III polyketide synthases from Citrus microcarpa. *J Biol Chem* **288**, 28845-58.

Morita H, Kondo S, Oguro S, *et al.*, 2007. Structural insight into chain-length control and product specificity of pentaketide chromone synthase from Aloe arborescens. *Chem Biol* **14**, 359-69.

Morita H, Shimokawa Y, Tanio M, *et al.*, 2010a. A structure-based mechanism for benzalacetone synthase from Rheum palmatum. *Proc Natl Acad Sci U S A* **107**, 669-73.

Morita H, Tanio M, Kondo S, *et al.*, 2008. Crystallization and preliminary crystallographic analysis of a plant type III polyketide synthase that produces benzalacetone. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **64**, 304-6.

Morita H, Wanibuchi K, Nii H, Kato R, Sugio S, Abe I, 2010b. Structural basis for the one-pot formation of the diarylheptanoid scaffold by curcuminoid synthase from Oryza sativa. *Proc Natl Acad Sci U S A* **107**, 19778-83.

Nualkaew N, Morita H, Shimokawa Y, *et al.*, 2012. Benzophenone synthase from Garcinia mangostana L. pericarps. *Phytochemistry* **77**, 60-9.

Raeth T, 2007. *Charakterisierung und Kristallisierung der Biphenylsynthase I von Sorbus aucuparia.*

Resmi MS, Verma P, Gokhale RS, Soniya EV, 2013. Identification and characterization of a type III polyketide synthase involved in quinolone alkaloid biosynthesis from Aegle marmelos Correa. *J Biol Chem* **288**, 7271-81.

Revill WP, Bibb MJ, Hopwood DA, 1996. Relationships between fatty acid and polyketide synthases from Streptomyces coelicolor A3(2): characterization of the fatty acid synthase acyl carrier protein. *J Bacteriol* **178**, 5660-7.

Romero-Perez AI, Ibern-Gomez M, Lamuela-Raventos RM, De La Torre-Boronat MC, 1999. Piceid, the major resveratrol derivative in grape juices. *J Agric Food Chem* **47**, 1533-6.

Rosaria Perrone PDR, Olga De Castro,Paolo Colombo, 2013. Leaf and stem anatomy in eight Hypericum species (Clusiaceae). *Acta Botanica Croatica* **72**, 269-86.

Russo Krauss I, Merlino A, Vergara A, Sica F, 2013. An overview of biological macromolecule crystallization. *Int J Mol Sci* **14**, 11643-91.

Sankaranarayanan R, Saxena P, Marathe UB, Gokhale RS, Shanmugam VM, Rukmini R, 2004. A novel tunnel in mycobacterial type III polyketide synthase reveals the structural basis for generating diverse metabolites. *Nat Struct Mol Biol* **11**, 894-900.

Sanmartín ASMI, 2012. Paleobiology of the genus Hypericum(Hypericaceae): a survey of the fossil record and its palaeogeographic implications. *Anales del Jardín Botánico de Madrid* **69(1): 97-106 .**

Satou R, Miyanaga A, Ozawa H, *et al.*, 2013. Structural basis for cyclization specificity of two Azotobacter type III polyketide synthases: a single amino acid substitution reverses their cyclization specificity. *J Biol Chem* **288**, 34146-57.

Saxena P, Yadav G, Mohanty D, Gokhale RS, 2003. A new family of type III polyketide synthases in Mycobacterium tuberculosis. *J Biol Chem* **278**, 44780-90.

Shen B, 2003. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr Opin Chem Biol* **7**, 285-95.

Shimokawa Y, Morita H, Abe I, 2010. Structure-based engineering of benzalacetone synthase. *Bioorg Med Chem Lett* **20**, 5099-103.

Shomura Y, Torayama I, Suh DY, *et al.*, 2005. Crystal structure of stilbene synthase from Arachis hypogaea. *Proteins* **60**, 803-6.

Smith S, Tsai SC, 2007. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat Prod Rep* **24**, 1041-72.

Springob K, Lukacin R, Ernwein C, Groning I, Matern U, 2000. Specificities of functionally expressed chalcone and acridone synthases from Ruta graveolens. *Eur J Biochem* **267**, 6552-9.

Staunton J, Weissman KJ, 2001. Polyketide biosynthesis: a millennium review. *Nat Prod Rep* **18**, 380-416.

Stevens P, 2007. *The families and genera of vascular plants*. springer berlin Heidelberg.

Stewart C, Jr., **Woods K**, **Macias G**, **Allan AC**, **Hellens RP**, **Noel JP**, 2017. Molecular architectures of benzoic acid-specific type III polyketide synthases. *Acta Crystallogr D Struct Biol* **73**, 1007-19.

Suh DY, Fukuma K, Kagami J, *et al.*, 2000. Identification of amino acid residues important in the cyclization reactions of chalcone and stilbene synthases. *Biochem J* **350 Pt 1**, 229-35.

Suhren O, 1951. [First clinical experience with pikromycin, a new antibiotic from actinomyces in therapy of pyoderma]. *Med Klin* **46**, 722-3.

Takasaki M, Konoshima T, Shingu T, *et al.*, 1990. Structures of euglobal-G1, -G2, and -G3 from Eucalyptus grandis, three new inhibitors of Epstein-Barr virus activation. *Chem Pharm Bull (Tokyo)* **38**, 1444-6.

Wanibuchi K, Morita H, Noguchi H, Abe I, 2011. Enzymatic formation of an aromatic dodecaketide by engineered plant polyketide synthase. *Bioorg Med Chem Lett* **21**, 2083-6.

Watanabe A, Ebizuka Y, 2004. Unprecedented mechanism of chain length determination in fungal aromatic polyketide synthases. *Chem Biol* **11**, 1101-6.

Xiao ZY, Mu Q, Shiu WK, Zeng YH, Gibbons S, 2007. Polyisoprenylated benzoylphloroglucinol derivatives from Hypericum sampsonii. *J Nat Prod* **70**, 1779-82.

Yamaguchi T, Kurosaki F, Suh DY, *et al.*, 1999. Cross-reaction of chalcone synthase and stilbene synthase overexpressed in Escherichia coli. *FEBS Lett* **460**, 457-61.

Yamakoshi Y, Murata M, Shimizu A, Homma S, 1992. Isolation and characterization of macrocarpals B--G antibacterial compounds from Eucalyptus macrocarpa. *Biosci Biotechnol Biochem* **56**, 1570-6.

Yin S, Xue JJ, Fan CQ, Miao ZH, Ding J, Yue JM, 2007. Eucalyptals A-C with a new skeleton isolated from Eucalyptus globulus. *Org Lett* **9**, 5549-52.

Yu D, Xu F, Zeng J, Zhan J, 2012. Type III polyketide synthases in natural product biosynthesis. *IUBMB Life* **64**, 285-95.

VIII. Appendix

1. Sequences of used enzymes

Hypericum sampsonii Benzophenone synthase (HsBPS)

ATGGCCCCTGCAATGGAGTACTCAACCCAGAACGGCCAGGGGGAGGGAAAGAAGAGGGCCAGTGTCCTCG
CCATTGGAACGACCAACCCGGAGCACTTCATCTTGCAGGAGGACTACCCGGACTTCTACTTCAGGAACACC
AACAGCGAGCACATGACCGATCTCAAGGAGAAATTCAAGCGCATCTGTGTGAAGTCTCATATTAGGAAGA
GGCACTTCTACCTGACCGAGGAGATCCTCAAGGAGAACCAGGGCATCGCCACCTACGGCGCGGGCTCCCT
GGACGCCCGCCAGAGGATCCTTGAGACTGAGGTCCCCAAGCTAGGCCAAGAGGCGGCCCTCAAGGCCATC
GCGGAGTGGGGCCAGCCCATCTCCAAGATCACCCACGTGGTGTTTGCGACCACCTCCGGGTTCATGATGCC
CGGCGCAGACTACGCCATCACCCGCCTCcTCGGCCTCAACCGCACTGTTAGGCGCGTGATGCTCTACAACC
AGGGCTGCTTTGCCGGGGGCACGGCCCTCCGAGTCGCCAAGGACCTCGCAGAGAACAACGAAGGCGCGCG
GGTCCTCGTCGTGTGCGCGGAGAACACCGCCATGACTTTCCACGCTCCCAACGAGTCCCACCTCGACGTGA
TCGTGGGGCAGGCCATGTTCTCAGATGGTGCTGCCGCTCTGATCATCGGGGCAGGCCCTGACGTTGCCGCG
GGGGAGCGCGCAGTGTTCAATATCCTATCGGCGAGCCAGACGATCGTGCCGGGCTCCGACGGGGCGATAA
CGGCGCACTTCTACGAGATGGGGATGAGCTACTTCCTTAAGGAGGACGTCATCCCTCTCTTCAGGGACAAC
ATCGCCGCCGTCATGGAGGAGGCCTTCTCCCCGCTTGGGGTCTCCGACTGGAACTCCCTCTTCTACTCCATC
CACCCCGGCGGCCGCGGTATCATCGACGGAGTCGCCGGGAACCTCGGGATCAAGGACGAGAACCTCGTGG
CCACCAGGCACGTCCTCGGCGAGTACGGGAACATGGGGTCAGCGTGCGTCATGTTCATCCTGGACGAGCTC
AGGAAGAGCTCCAAGCTCAGCGGGAAGCCCACCACCGGCGACGGCAAAGAGTTCGGCTGCCTCATCGGCC
TCGGCCCCGGCCTCACCGTGGAGGCCGTTGTCCTCCAGAGTGTCCCCATCCTCCAGTGA

MAPAMEYSTQNGQGEGKKRASVLAIGTTNPEHFILQEDYPDFYFRNTNSEHMTDLKEKFKRICVKSHIRKRHFY
LTEEILKENQGIATYGAGSLDARQRILETEVPKLGQEAALKAIAEWGQPISKITHVVFATTSGFMMPGADYAITR
LLGLNRTVRRVMLYNQGCFAGGTALRVAKDLAENNEGARVLVVCAENTAMTFHAPNESHLDVIVGQAMFSD
GAAALIIGAGPDVAAGERAVFNILSASQTIVPGSDGAITAHFYEMGMSYFLKEDVIPLFRDNIAAVMEEAFSPLG
VSDWNSLFYSIHPGGRGIIDGVAGNLGIKDENLVATRHVLGEYGNMGSACVMFILDELRKSSKLSGKPTTGDGK
EFGCLIGLGPGLTVEAVVLQSVPILQ

Hypericum androsaemum Benzophenone synthase (HaBPS)

ATGGCCCCGGCGATGGAGTACTCAACCCAGAACGGCCAGGGGGAGGGAAAGAAGAGAGCTAGTGTCCTCG
CTATTGGAACAACCAACCCGGAACATTTCATCTTGCAGGAGGACTACCCCGACTTCTACTTCAGGAACACC
AACAGCGAGCACATGACCGAGCTCAAGGAGAAATTTAAACGTATCTGTGTTAAGTCTCATATTAGGAAGA
GGCACTTCTACCTAACCGAGGAGATTCTCAAGGAGAACCAGGGGATCGCCACCTATGGCGCGGGCTCCCTT
GACGCCCGCCAGAGGATCCTCGAGACCGAGGTCCCGAAGCTGGGTCAGGAGGCGGCCCTCAAAGCCATCG
CGGAGTGGGGCCAGCCCATCTCCAAGATAACACACGTGGTGTTCGCGACGACCTCCGGGTTCATGATGCCC
GGGGCAGACTACGTCATCACCCGCCTCCTCGGCCTCAACCGCACCGTCAGGCGCGTCATGCTCTACAACCA
GGGCTGCTTCGCTGGGGGCACGGCCCTCCGTGTCGCCAAGGACCTCGCGGAGAACAACGAGGGCGCGCGC
GTGCTCGTCGTGTGCGCGGAGAACACCGCCATGACTTTCCACGCCCCCAACGAGTCCCACCTAGACGTGAT
CGTGGGGCAAGCCATGTTCTCAGATGGCGCGGCTGCTCTGATCATCGGGGCATGCCCTGACGTTGCTTCTG
GGGAGCGCGCAGTGTTCAATATCCTGTCGGCGAGCCAGACGATCGTGCCGGGCTCCGACGGGGCGATAAC
GGCGCACTTCTACGAGATGGGGATGAGCTACTTCCTTAAGGAGGACGTCATCCCTCTCTTCCGTGATAACA
TCGCCGCCGTCATGGAGGAGGCCTTCTCTCCGCTTGGGGTCTCCGACTGGAACTCCCTCTTCTACTCCATCC
ACCCCGGTGGCCGAGGGATCATCGACGGCGTCGCCGGGAACCTTGGGATCAAGGACGAGAACCTTGTGGC
GACCAGGCACGTCCTCGGCGAGTACGGGAACATGGGGTCAGCCTGCGTGATGTTCATCCTTGACGAGCTTA
GGAAGAGCTCCAAGGTCAACGGGAAGCCCACCACCGGCGACGGCAAGGAGTTCGGCTGCCTCATCGGCCT
CGGCCCTGGCCTCACCGTGGAAGCCGTCGTCCTCCAGAGTGTCCCAATTCTCCAGTGA

MAPAMEYSTQNGQGEGKKRASVLAIGTTNPEHFILQEDYPDFYFRNTNSEHMTELKEKFKRICVKSHIRKRHFYLTEEILK
ENQGIATYGAGSLDARQRILETEVPKLGQEAALKAIAEWGQPISKITHVVFATTSGFMMPGADYVITRLLGLNRTVRRVML
YNQGCFAGGTALRVAKDLAENNEGARVLVVCAENTAMTFHAPNESHLDVIVGQAMFSDGAAALIIGACPDVASGERAVFNI
LSASQTIVPGSDGAITAHFYEMGMSYFLKEDVIPLFRDNIAAVMEEAFSPLGVSDWNSLFYSIHPGGRGIIDGVAGNLGIK
DENLVATRHVLGEYGNMGSACVMFILDELRKSSKVNGKPTTGDGKEFGCLIGLGPGLTVEAVVLQSVPILQ

Malus domestica biphenyl synthase 3 (MdBIS3)

ATGGCGCCTTTGGTTAAGAATGAGCCTCAACATGCCAAAATCCTAGCCATTGGCACTGCAAATCCACCCAA
CGTCTACCACCAAAAGGATTATCCTGATTTCTTGTTTCGAGTTACCAAAAATGAGCACAGAACAGATTTAA
GAGAGAAGTTCGATCGCATTTGTGAGAAATCAAGAACAAAGAAGCGCTACTTGCATCTAACAGAAGAGAT
GTTGAAAGCTAACCCAAACATATACACCTATGGAGCCCCATCACTCGATGTGCGCCAAGACATTTGTAACA
TTGAGGTCCCAAAGCTAGGGCAAGAAGCAGCATTGAAAGCCATCAAAGAGTGGGGCCAGCCCATTTCAAA
AATCACCCACCTCATCTTTTGCACAGCTTCTTGCGTTGACATGCCAGGTTGTGACTTCCAATTAATCAAGCT
CCTCGGCCTTGATCCATCCGTCACCAGAACCATGATCTATGAAGCTGGCTGCTATGCTGGTGCGACAGTCCT
CCGCATGGCCAAAGACTTCGCAGAGAACAATAAGGGCGCACGAGTCCTTGTGGTGTGCGCCGAGATCACG
ACCGTGTTTTTCCACGGACTCACTGACACCCACCTTGACATATTGGTGGGTCAAGCTCTTTTTGCTGACGGA
GCATCTGCTGTGATAGTTGGGGCCAATCCAGAGCCTGAAATTGAGAGGCCACTGTTTGAAATCGTGGCTTG
TAGGCAAACGATCCTACCAAACTCAGAGCATGGTGTGGTGGCCAACATTCGTGAAATGGGGTTTAATTATT
ATTTATCAGGAGATGTCCCCAAATTCGTTGGTGGAAATGTTGTGGATTTTATGACTAAAACTTTTGAAAAG
GTAGATGGGAAGAAAAAGGACTGGAACTCCTTGTTTTTCAGTGTGCACCCTGGTGGACCTGCCATTGTAGA
CCAGGTGGAGGAGAAATTGGGTTTGAAGGAAGGGAAGCTTAGGGCAACAAGGCATGTGTTGAGTGAGTAT
GGCAACATGGGAGCTCCAACTGTGCACTTTATTTTGGATGAGATGAGAAATAAGTCGATTGAGGAAGGCA
AAACCACTACTGGTGAAGGTTTGGAATGGGGTGTCGTGATTGGAATCGGACCGGGACTCACTGTCGAGAC
AGCCGTGCTGCGTAGTGAATCTATTCCATGCTAA

MAPLVKNEPQHAKILAIGTANPPNVYHQKDYPDFLFRVTKNEHRTDLREKFDRICEKSRTKKRYLHLTEEMLKANPNIYTY
GAPSLDVRQDICNIEVPKLGQEAALKAIKEWGQPISKITHLIFCTASCVDMPGCDFQLIKLLGLDPSVTRTMIYEAGCYAG
ATVLRMAKDFAENNKGARVLVVCAEITTVFFHGLTDTHLDILVGQALFADGASAVIVGANPEPEIERPLFEIVACRQTILP
NSEHGVVANIREMGFNYYLSGDVPKFVGGNVVDFMTKTFEKVDGKKKDWNSLFFSVHPGGPAIVDQVEEKLGLKEGKLRAT
RHVLSEYGNMGAPTVHFILDEMRNKSIEEGKTTTGEGLEWGVVIGIGPGLTVETAVLRSESIPC

Sorbus aucuparia biphenyl synthase 1 (SaBIS1)

ATGGCGCCTTTGGTTAAGAATCATGGAGAGCCTCAACATGCCAAAATCCTAGCCATTGGCACTGCAAATCC
ACCAAACGTCTACTACCAAAAAGACTATCCTGATTTCTTGTTTCGAGTCACCAAAAATGAGCACAGGACAG
ATTTAAGAGAGAAGTTTGATCGCATTTGTGAGAAATCAAGAACAAGGAAGCGTTACTTGCATCTAACAGA
GGAGATTCTAAAGGCTAACCCAAGCATATATACCTATGGTGCCCCATCACTCGATGTGCGCCAAGACATGT
TGAATTCTGAGGTCCCAAAGCTAGGGCAACAAGCAGCACTGAAAGCCATCAAAGAGTGGGGCCAACCCAT
CTCAAAGATCACCCACCTCATCTTTTGCACAGCTTCATGTGTTGACATGCCAGGTGCCGACTTCCAATTGGT
CAAGCTCCTCGGCCTTAACCCATCTGTCACTAGAACCATGATCTACGAAGCTGGTTGCTATGCTGGTGCAA
CTGTCCTCCGCCTAGCCAAGGACTTCGCAGAGAACAACGAGGGTGCACGCGTCCTTGTGGTGTGCGCTGAG
ATCACGACCGTGTTCTTCCACGGACTCACTGACACCCACTTGGACATACTGGTGGGCCAGGCTCTTTTTGCT
GACGGAGCATCTGCTGTGATAGTTGGGGCCAATCCAGAGCCTAAAATTGAGAGGCCACTATTTGAAATCGT
GGCATGCAGGCAGACAATCATACCGAACTCAGAGCATGGTGTGGTGGCCAACATTCGTGAAATGGGGTTT
ACTTATTATTTATCAGGAGAAGTCCCCAAATTTGTTGGTGGAAATGTTGTGGATTTTCTGACTAAAACTTTT
GAAAAAGTTGACGGAAAGAATAAGGACTGGAACTCCTTGTTTTTCAGTGTGCACCCTGGTGGACCCGCCAT
TGTAGACCAGGTGGAGGAGCAACTGGGTTTGAAGGAAGGGAAGCTTAGGGCAACAAGGCATGTGTTGAGT
GAGTATGGCAACATGGGAGCTCCATCTGTGCACTTTATTTTGGATGATATGAGAAAGAAGTCGATTGAGGA
AGGCAAATCCACAACTGGCGAAGGTTTGGAATGGGGTGTCGTGATTGGAATCGGACCAGGACTCACTGTT
GAGACAGCCGTACTGCGTAGTGAATCTATTCCATGCTAAATACTTGCATAAAATGATTCATATATTGATTA

MAPLVKNHGEPQHAKILAIGTANPPNVYYQKDYPDFLFRVTKNEHRTDLREKFDRICEKSRTRKRYLHLTEEILKANPSIY
TYGAPSLDVRQDMLNSEVPKLGQQAALKAIKEWGQPISKITHLIFCTASCVDMPGADFQLVKLLGLNPSVTRTMIYEAGCY
AGATVLRLAKDFAENNEGARVLVVCAEITTVFFHGLTDTHLDILVGQALFADGASAVIVGANPEPKIERPLFEIVACRQTI
IPNSEHGVVANIREMGFTYYLSGEVPKFVGGNVVDFLTKTFEKVDGKNKDWNSLFFSVHPGGPAIVDQVEEQLGLKEGKLR
ATRHVLSEYGNMGAPSVHFILDDMRKKSIEEGKSTTGEGLEWGVVIGIGPGLTVETAVLRSESIPC

2. Crystallization kits

Table 9. The crystallization screens used in this project

QIAGEN	Molecular Dimensions	Hampton research	Jena bioscience
JCSG1 JCSG2 JCSG3 JCSG4 JCSG+ PACT	PGA MORPHEUS MIDAS+	INDEX	JBScreen ClassicHTS2

3. Synchrotron trips

Hamburg: Deutsches Elektronen-Synchrotron (DESY) in Helmholtz-Gemeinschaft.

Switzerland: Swiss Lightsource (SLS) at the Paul Scherrer Institut (PSI).

Grenoble-France: The High Throughput Crystallization Facility at HTX lab using Crystallization Information Management System (CRIMS).

4. PDB-code of previously mentioned crystallized plant type III PKSs

Table 10. The PDB numbers of crystallized plant type III PKSs mentioned in this project

Enzyme	PDB-code
MsCHS	1BI5
PsSTS and AhSTS	1XES and 1Z1E
GhPS	1QLV
CClCUS	3OV2
OsCUS	3OIT
RpBAS	3A5Q
CmACS	3WD7
CmQNS	3WD8
AaPCS	2D3M
AaPCSM207G (function of OKS)	2D51
HaBPS	5UCO
MdBIS3	5W8Q
MdBIS3 complexed with benzoyl-CoA	5WC4

5. Expression, purification and crystallization of the mutant HsBPST135K

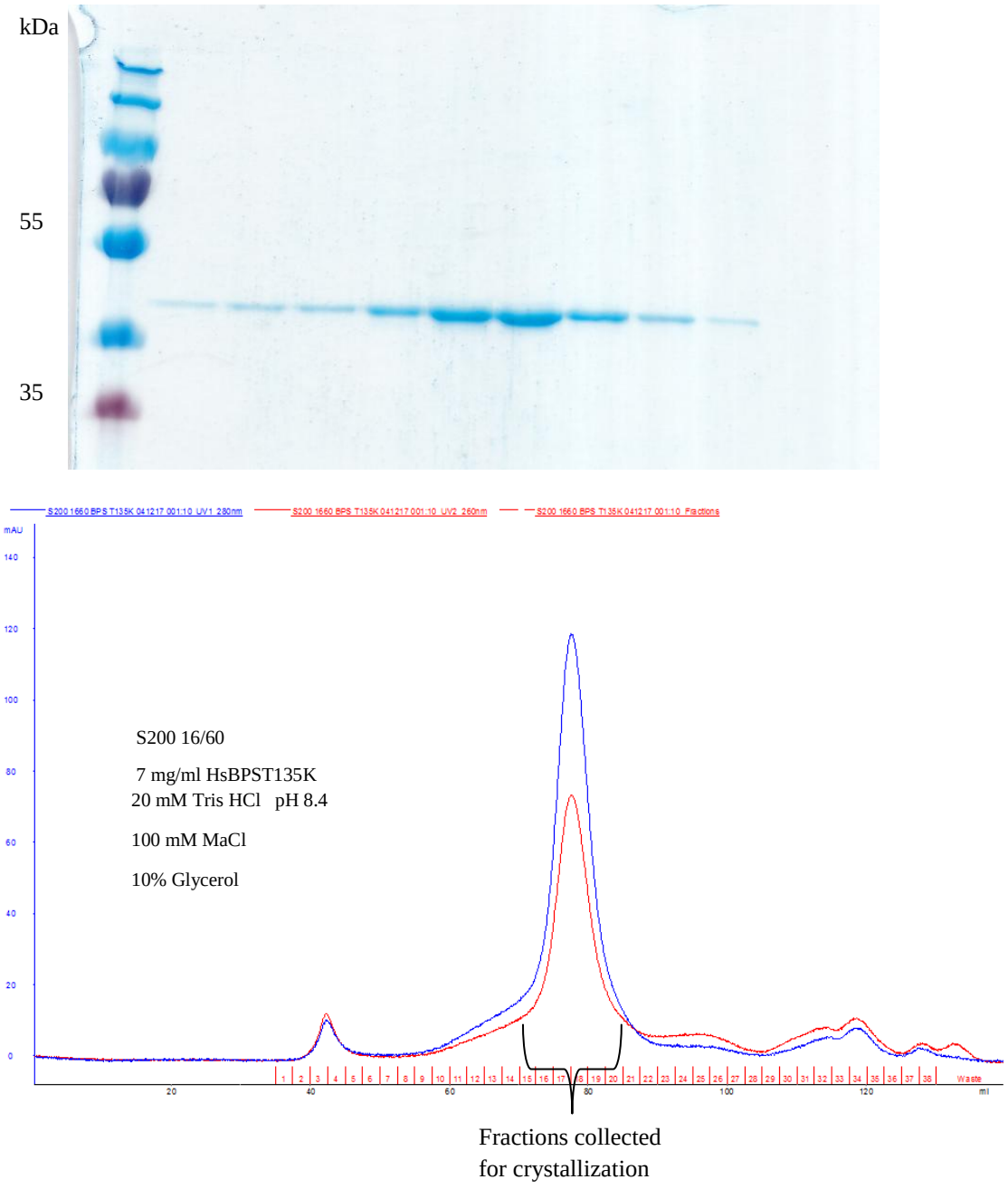


Fig VIII.1a First purification trial of HsBPST135K. SDS-PAGE shows pure protein and a relatively sharp peak appeared in gel filtration chromatography. The specific fractions from 16 to 21, which correspond to the main sharp peak, were collected, concentrated and subjected to crystallization screens.

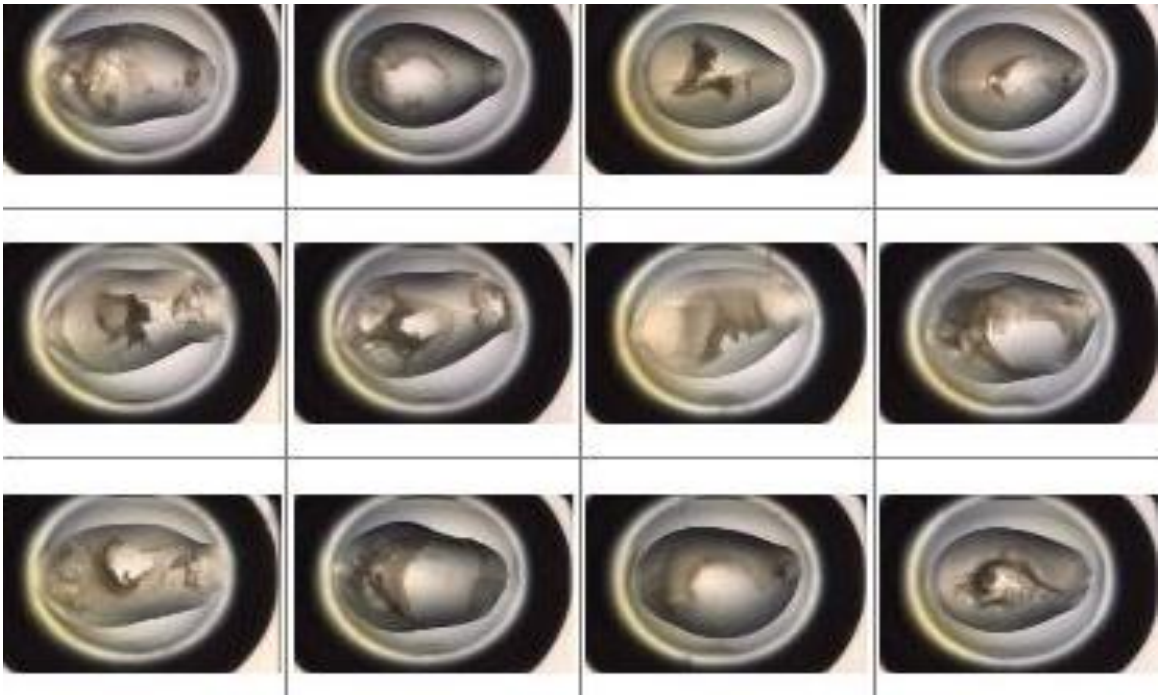


Fig VIII.1b The first crystallization screens of HsBPST135K. The mutant did not develope any crystals, but it showed heavy precipitation in 95% of conditions after 12 h.

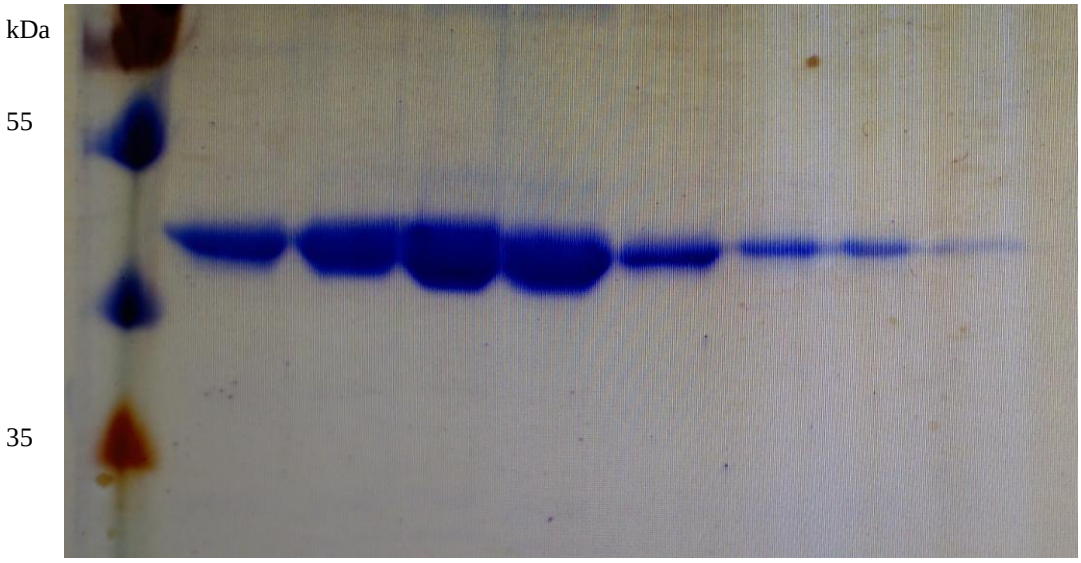


Fig VIII.2a Second purification trial of HsBPST135K. Large culture volumes were used to increase the protein amount more than in the first trial. SDS-PAGE shows highly concentrated pure protein.

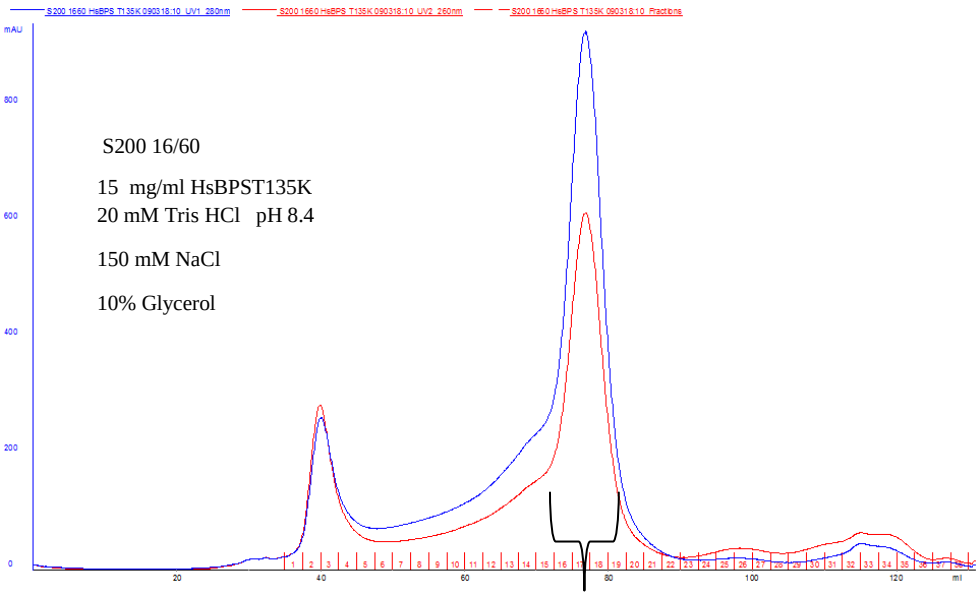


Fig VIII.2b A relatively sharp peak appeared in gel filtration chromatography of HsBPST135K. The specific fractions from 16 to 21, which correspond to the main sharp peak, were collected, concentrated and subjected to crystallization screens.

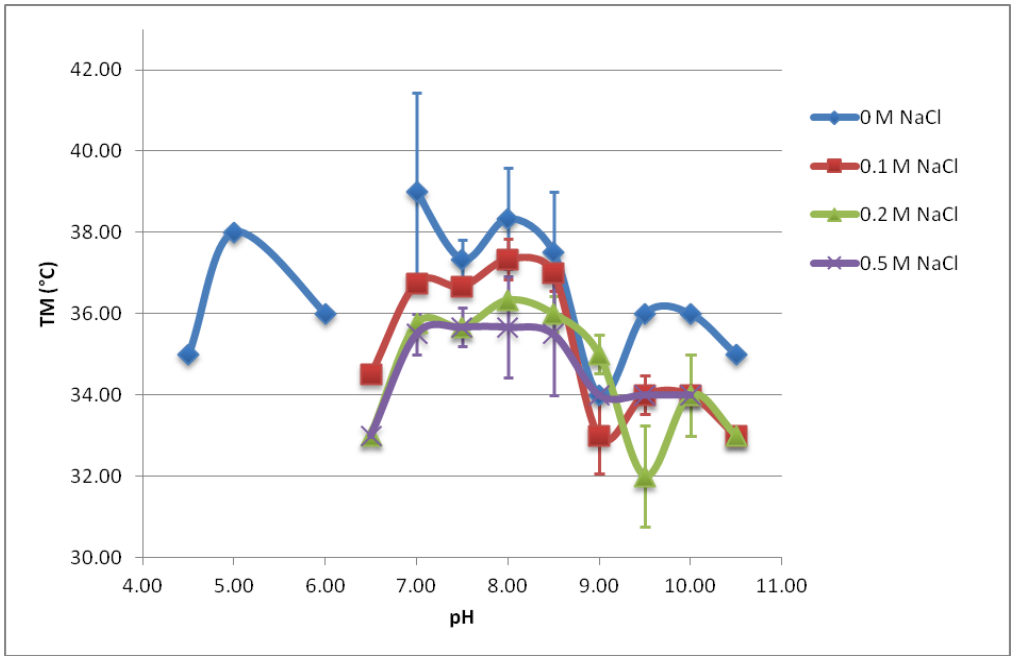


Fig VIII.3a Thermofluor assay results to detect the best buffer for HsBPST135K. High amounts of NaCl (100-500 mM) resulted in low stability of the protein in Tris-HCl buffer pH 8.5.

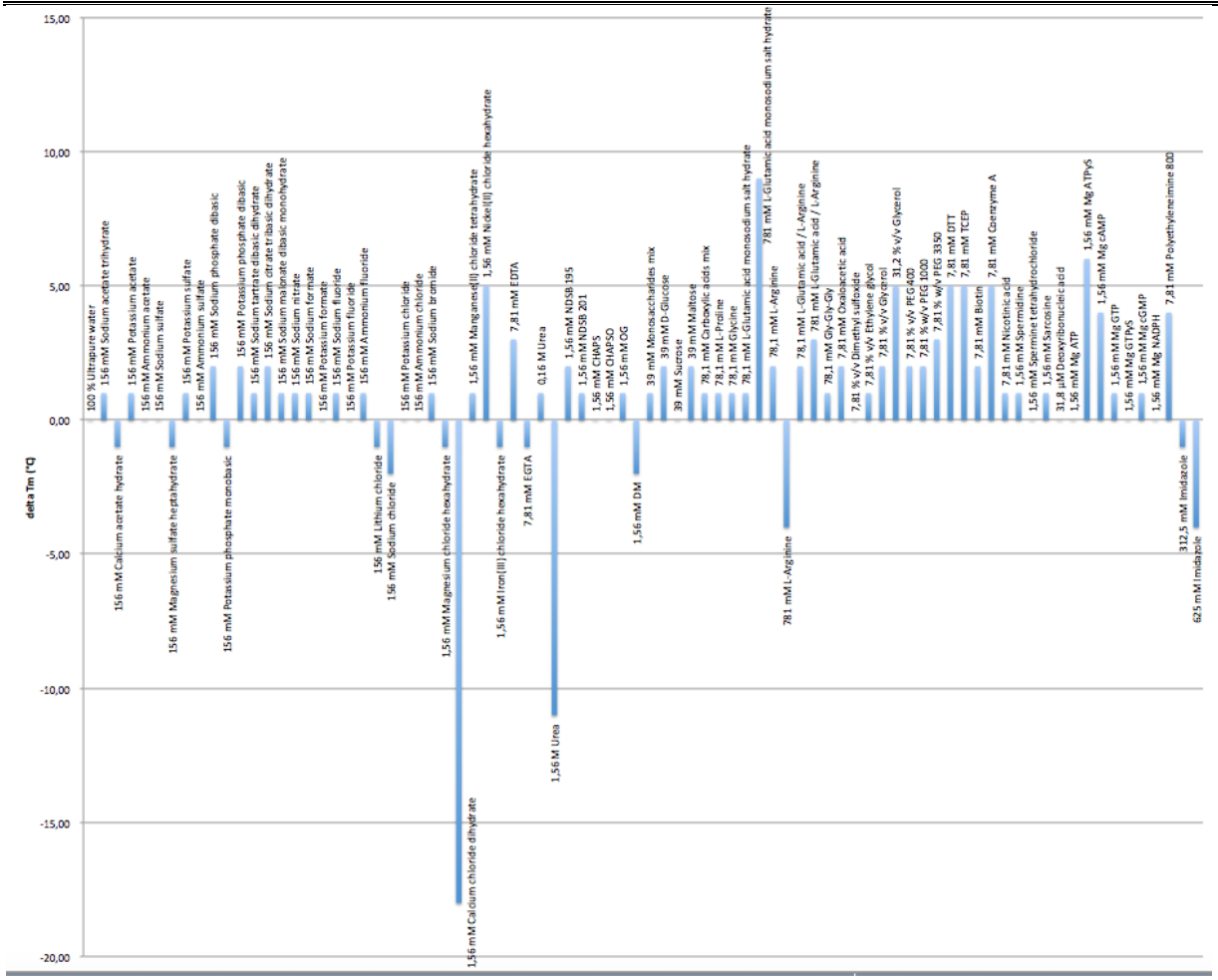


Fig VIII.3b Thermoflour assay results to detect the best additive for HsBPST135K. High thermal signals were recorded, but they did not show better stability of HsBPST135K in the buffer.

17 mg/ml protein in 20 mM Tris-HCl buffer was used in crystallization screens, but again heavey precipitate formed in almost all conditions.